Infrared Array Spectroscopy with Cooled Grating Spectrometer 4

Suzanne Katherine Ramsay

Presented for the Degree of Doctor of Philosophy
at the University of Edinburgh
1993

This thesis is my own
composition, except where
specifically indicated in the text.

March 1993.

Dedicated to my parents,

William and Anne Ramsay.

Abstract

The work contained in this thesis is concerned with the design and construction of a near–infrared spectrometer for astronomical applications, Cooled Grating Spectrometer 4 (CGS4). The results of two observational projects carried out with the instrument on the United Kingdom Infrared Telescope are presented.

CGS4 is one of the first infrared spectrometers to incorporate a 2D detector array, and the design of the spectrometer is driven by the desire to maximise the improvements in sensitivity which can be obtained with such an array. The need for high throughput and good image quality is discussed, and the way in which they have been achieved outlined. Other factors which affect the sensitivity of the instrument are the sky–background emission, thermal background from the instrument and telescope and detector noise. The method by which each of these is minimised is outlined. CGS4 was designed as an instrument which could be applied to the disparate projects which benefit from observations in the NIR waveband range. These include observations of molecules in star–forming regions, studies of emission lines from active galactic nuclei, and studies of gas dynamics.

Two aspects of optimising observations in the NIR are discussed in detail. For observations in the 1–2.3μm region, emission from hydroxyl in the Earth's atmosphere dominates the sky background. The intensity of the line emission from OH varies by $\sim$10% on timescales of $\sim$10mins. As a result, on–source exposure times should be restricted to 60s, setting a fundamental limit to the sensitivity achievable for observations of extended sources. Flat–fielding a 2D array spectrometer, to calibrate the relative gains of the pixels, is another area which is explored in this work. The reasons for excluding the methods used for optical astronomy and NIR imaging are discussed. The solution for CGS4 was to provide with a tailor–made "calibration unit". The design of this unit and the resulting success in flat–fielding observations is detailed.

The astronomical results reported concern observations of emission from the hydrogen molecule. There are two predominant methods of exciting H_2 : radiative excitation by UV photons with energies less than 13.6eV or collisional excitation in shocks. Excited H_2 decays by transitions in the rotational–vibrational bands of the ground electronic state, emitting NIR photons. The emission spectrum, specifically the 1–0 S(1)/2–1 S(1) line ratio, is frequently used to diagnose the excitation mechanism. Recent models of the emission from H_2 have shown that for gas densities above a critical density, the emission from radiatively excited gas can emulate that from a collisionally excited gas, and that radiative excitation of H_2 may be more widespread than previously thought.

The planetary nebula, Hubble 12, was identified by Dinerstein *et al.* (1987) as a source of fluorescent emission from H_2 . A spectrum taken with CGS4 confirms this result, in the light of the recent models, through observations of the emission from higher rotational–vibrational levels not previously detected in this source. The emission is shown to arise from a gas of density 10^4–$10^5 cm^{-3}$ illuminated by a source of UV field strength 10^4 times that of the interstellar medium.

H_2 emission from a ring of molecular material which surrounds the Galactic centre

was identified as being shock excited when first measured by Gatley *et al.* (1984). Conditions at the Galactic centre, and the evidence for recent star–formation, have prompted a re–measurement of the H_2 spectrum from this source. It is shown that, indeed, radiative excitation of a dense ($\sim 10^6 cm^{-3}$) gas by a UV field of $10^5 G_o$ is sufficient to explain the H_2 observations. The dynamics of the gas has also been investigated. The observations confirm that the ring has a radius of 1.54pc, at which point the velocity of the gas is 100km s^{-1} , and that the velocity has a keplerian dependence on radius. This is consistent with the gas orbiting in the gravitational field of a compact, central source of 4×10^6 $M_\odot$.

Acknowledgements

As the holder of an SERC postgraduate studentship, I acknowledge their support for the duration of my studies and their funding of my trips to foreign parts.

During my three and a half years with the Department of Astronomy, I have luxuriated under a higher than average supervisor–student ratio, and it is with great pleasure that I thank each of Matt Mountain, David Robertson, Peter Brand, Terry Lee and John Cooke for their advice, support and assistance during this period. I am particularly indebted to Matt, with whom I have worked most closely. He has been a good friend and supporter in my endeavours to produce this thesis. Many thanks are also due to David Robertson, for patient instruction in the ways of engineering.

As a member of the CGS4 project team, it was my great pleasure to work with everyone else involved in the instrument. David Montgomery and Gordon Pentland were always ready to answer stupid questions, and my brief course in grit–blasting from Dave remains the highlight of my first year at ROE. Rosemary Glendinning taught me all I know about the CGS4 electronics, and we're still friends! On first arriving, I was made to feel welcome by John Clark and Ron Beetles in particular, and I am grateful to them for that. Of the others in the CGS4 team, I would like to acknowledge Davie, Joel, Eli, Alan, Len and Steven for all their help.

There are many other people whose help in the progress of my studies I would like to mention here. Notable amongst them are Phils Blanco, Puxley and James, especially the latter pair who have taken the time to read this thesis. It is always a pleasure to talk "optics" with Alistair Glasse, although I'm not sure I always get the joke, Alistair? I've had much reason to be glad that Stuart knows everthing, and thanks also to Horst. I have enjoyed working with Tom Geballe in the past and hope to continue doing so. I am also looking forward to continuing a collaboration with Ian Gatley. Thanks to the UKIRT telescope operators, Dolores, Joel and Thor, for all their help, without which much of this thesis really wouldn't have happened. And to Thor, for the blues. The University equivalent of the TO must be Liz Gibson, who deserves a mention for doing it all! Maureen has alway been extremely helpful, and I don't know how to thank her for the last minute photocopying help!

I have been lucky enough to make a number of observing trips, and extended tours of instrumental work. My enjoyment of these trips has been greatly enhanced by the friendship, hospitality, and often accomodation, that I have received from the staff of the Joint Astronomy Centre and the Cerro Tololo Inter–American Observatory. I would particularly like to thank Brooke and Elizabeth Gregory, fellow Glaswegian Gillian Wright and Phil Puxley. And Matt and Rosemary for the jacuzzi. Also Bill Dent, Simon Craig, Sidney Arakaki, Vernon de Mattos, Ant and Ros, who have enlivened my trips to Hilo.

Meanwhile, back at the ranch. Many, many thanks go to Pippa, for her friendship, for lots of support and for the happy tradition of lavish dinners. Thanks to Antonio, who has been a great friend and source of endless CGS4 questions. To Dave and Amadeu, for being "our year". Dave and Martyn also get a special thank-you for use of their non-Starlink facilities. Thanks to Ruth for the jumping. I must also thank the Space

Shuttles: Ant, Ruth, Mike, Karen, Steven and Siobhan for many happy hours learning to be good losers, and for agreeing to the name. There have been many other people who have made a great difference to life at the Obsevatory, and I would like to mention Torch, Kathy, Isabelle, Bobs Nichol and Mann, Andys Conolly (for his alleged mood swings–quite an act to follow) and Taylor, Shuggie, Clare, Penny, Rijdat, Steven, Martyn and Lance. And Nobby, the LaTeXsheep.

Of non–observatory folk that deserve much thanks, I am grateful to Sarah, Susie, Tom, Siobhan and Lyndsay, for their continuing friendship. My flat–mate, Gordon, who has shown remarkable patience of late, is as good as they come, and I particularly thank him for the macaroni. Thanks to my sister, Louise, for an uncanny knack of knowing when you need to hear from her. Lastly, to my long–suffering Himself, David, who knows more about East–West relations than Gorbachev, and for whom "normality" is (always) just round the corner. I owe him a lot of love and support, and look forward to repaying it.

It is with great pride (and tremendous relief!) that I dedicate this thesis to my parents. To my mother, whose unwavering enthusiasm has been of immeasurable value, and to my father, who, despite my best efforts, still knows more physics than me!

Contents

Chapter 1

Introduction

In his "Experiments on the Refrangibility of the Invisible Rays of the Sun" William Herschel (1800) relates his discovery of infrared radiation. During his initial experiments on this new phenomenon he noted that the rays, 'though not fit for vision, were yet highly invested with a power for occasioning heat' and supposed that 'if we call *light* those rays which illuminate objects, and *radiant heat*, those which heat bodies, it may be inquired whether light be essentially different from radiant heat?'. His continued work on these 'rays which occasion heat' proved that they obeyed the same rules of refraction and reflection as did visible light. This paper marks the beginnings of a new branch of science, that of infrared physics. The work of this thesis is concerned with a new spectrometer for astronomical applications in the infrared, CGS4, and the impact of this new instrument on results attainable with near infrared (1–5μm ; NIR) observations. In this introductory Chapter, the motivation behind this work is explained in terms of the advantages of infrared astronomy, the gains to be achieved using a 2D detector array spectrometer and the astronomical projects which can be undertaken.

Cooled Grating Spectrometer 4 (CGS4) was commissioned on the United Kingdom Infrared Telescope (UKIRT) on Mauna Kea, Hawaii, during January to April 1991; first light was achieved on February 4th. CGS4 has been designed as a comprehensive instrument for NIR astronomy; it is one of the first spectrometers to take advantage of the recent availability of 2D detector arrays sensitive to NIR wavelengths and as

such represents a tremendous advance in this field. The 58×62 pixels of the CGS4 array give a thousandfold increase in the sensitivity of this instrument over the linear detectors of its predecessors, as well as providing a facility for broad–band, moderate resolution ($R=\frac{\lambda}{\Delta\lambda}$ ~400) spectroscopy and simultaneous coverage of $90\widehat{''}\times3\widehat{''}$ areas of spatially extended objects. It is not only due to the advances in detector technology that CGS4 has come about at this time; the latest developments in all areas from the mirror manufacture to cryogenic systems have contributed to the performance of CGS4 . The result is the most sensitive NIR spectrometer available on any telescope at the time of writing; a 3σ observation of a K~17 object can be made in 30min with low resolution grating (R~400). Details of the instrument design are given in Chapter 2. Chapter 3 describes in detail the testing and calibration of the instrument. The optical alignment of CGS4 was achieved using a static, mechanical procedure rather than standard alignment methods. This results in positional accuracies, for example of the mirrors and filters, within tolerances of 0.025mm decentre and 25mrad tilt, with aberrations in the image quality of less than 5% for off–axis images. The calibration of a 2D array spectrometer is another new area of endeavour; in particular, the difficulties associated with flat–fielding the array are significantly greater than those experienced with a camera or a visible light spectrometer. CGS4 uses a dedicated Calibration Unit containing spectral line sources for wavelength calibration and continuum sources to provide a flat–field. These features are described in Chapter 2, where a report is given of work carried out on the wavelength and temporal dependence of flat–fields and the flat–fielding accuracy which has been obtained.

Although NIR astronomy holds many advantages over other waveband regions, as discussed in Section 1.1, there are also pitfalls peculiar to this waveband. A source of background noise which remained largely ignored until the advent of staring arrays is the line emission from hydroxyl in the Earth's atmosphere. The ro–vibrational emission lines from this species are observed in the 1–2.3μm region with line strengths of thousands of photons per second, making this the dominant background emission in this wavelength region. Moreover, the emission from OH is variable with periods of ~10mins, and has been measured to vary by ~10% of its strength (Chapter 4). The strength of the emission lines are comparable with the strength of the H_2 lines observed at the Galactic centre,

and the signal from a 17th magnitude, extended source will be less than 1% of the OH line strength. This background affects NIR observations in many ways: exposure times are restricted to avoid gross variation of the OH line strength between object and sky exposures and the spectral structure obliges the infrared spectroscopist to invent novel flat–fielding schemes rather than adopting the standard technique of using sky observations. The noise on the residual signal from sky–subtracted spectra due the variation of the emission lines during the observation can provide a fundamental limit to the performance of staring arrays. Chapter 4 contains details of a study of this emission and its variability. Other sources of noise in the NIR are also covered in Chapter 2.

1.1 The Advantages of the Infrared Waveband Range

The motivation to design and build an infrared spectrometer comes from the advantages that an IR view of the universe gives the astronomer. This discussion is restricted to those features of the near infrared (1–5μm) which pertain to the astronomical work of this thesis (Chapters 5, 6), namely the greatly reduced extinction from interstellar dust at these wavelengths and the opportunities for observing molecular emission. This waveband is sub–divided into a number of atmospheric "windows" where almost all of the light from astronomical sources is transmitted to telescopes on the ground. A graph showing atmospheric transmission as a function of wavelength, with the windows labelled with their conventional names, is displayed in Figure 1.1; it has been generated using the IRTRANS model for the atmosphere (Traub & Stier 1976). The model parameters to be set are the amount of precipitable water in the atmosphere, taken to be 1.2mm (a value typical of the conditions on Mauna Kea) and the resolution, which was set to approximately that of the low dispersion grating in CGS4 (R=400). A triangular smoothing function was chosen to emulate the CGS4 line shape.

1.1.1 Obscured Regions and Reduced Extinction

The process of the scattering and absorption of starlight by dust grains in the interstellar medium is called extinction. It is expressed in terms of the intensity of light reaching

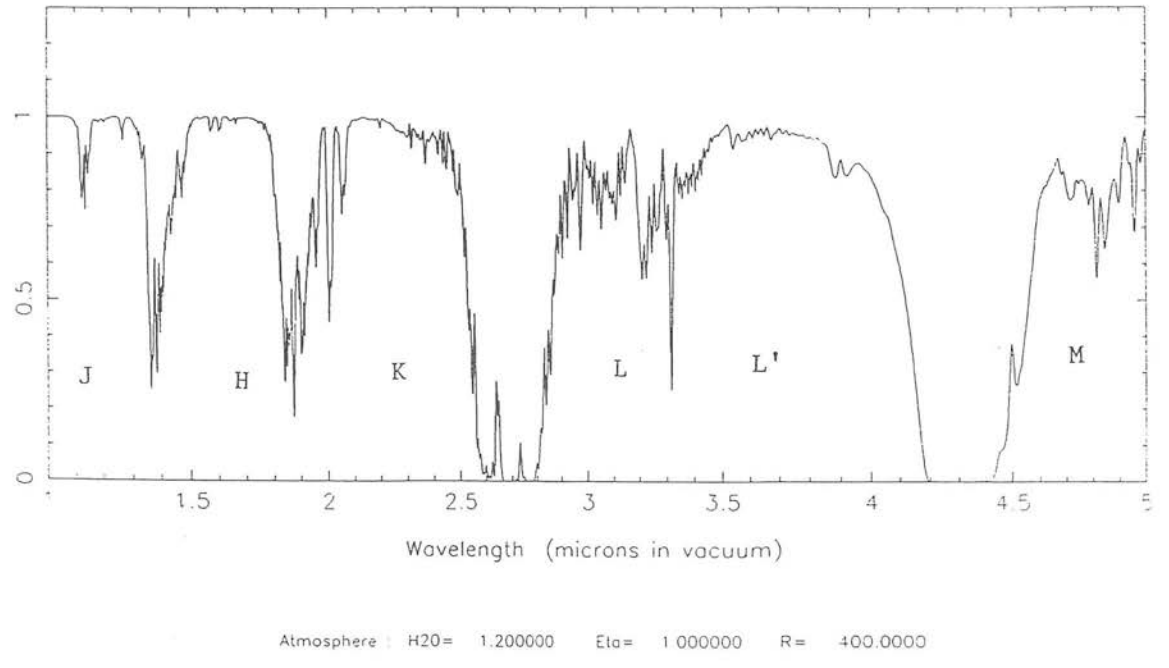


Figure 1.1: The transmission of the atmosphere, modelled for 1.2mm H_2O, a value typical of the atmosphere on Mauna Kea.

the observer (I_{obs}) compared with the observed intensity if the line of sight were dust free (I_o).

$$I_{obs} = I_o e^{-\tau_\lambda} \tag{1.1}$$

where τ_λ is referred to as the optical depth. The observed form of the extinction law for wavelengths from 0.002μm to 250μm is shown in Figure 1.2 (Mathis 1990); the vertical axis shows the extinction in magnitudes, A_λ ($= 1.086\tau_\lambda$) as a function of the visual extinction. It can be seen readily that the extinction at NIR wavelengths is greatly reduced compared with that at visible wavelengths. Thus the highly obscured regions such as star forming regions at the centre of molecular clouds can be observed in the infrared, making this the ideal waveband for studies of embedded sources such as protostars (see *e.g.* Rayner *et al.* 1990). One particular example of an object which cannot be viewed at visible wavelengths is the Galactic centre. The extinction to this object has been measured to be $A_v \sim 30$ (Wade *et al.* 1987), which reduces to $A_K \sim 3$. Observations of the H_2 emission from the central few parsecs of the Galaxy are presented in Chapter 6 of this thesis. Recently, Draine (1988) has provided a comprehensive study of the extinction law for the NIR, showing it to be a power law of the form $\lambda^{-1.75}$. The form of the extinction law is in itself a useful indicator of the nature of the dust grains.

1.1.2 Molecular Chemistry

Much of long wavelength astronomy is concerned with the study of molecules which are in rotational and vibrational motion: the energies of transitions between these levels are $\sim$1ev which corresponds to a wavelength of light of $\sim 1\mu$m . The astronomical observations contained in this thesis are of the simplest and most abundant molecule, hydrogen; descriptions of this molecule and the near infrared emission from it are contained in Sections 1.3 and 1.3.2. Examples of other molecular species observed in the NIR with

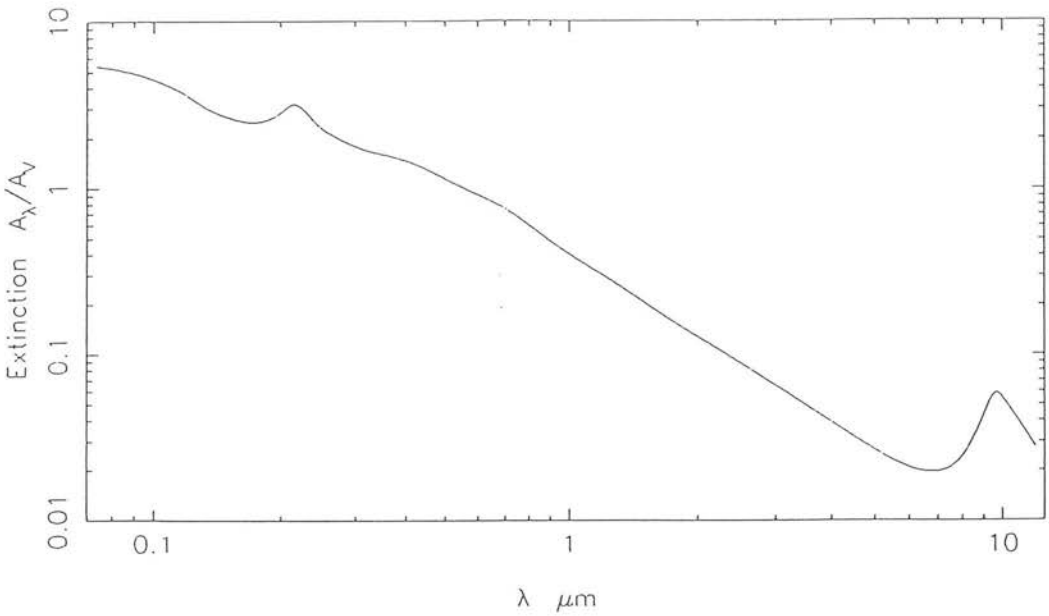


Figure 1.2: Extinction due to interstellar dust as a function of wavelength (Mathis 1990), illustrating the advantages of near infrared astronomy over optical astronomy for observations of obscured sources.

CGS4 include CO and more exotic molecules such as H_3^+. Recombination lines from many abundant species, including hydrogen and helium, fall in this waveband range.

1.2 The Advantages of Infrared Arrays

CGS4 is the successor to a linear array spectrometer, Cooled Grating Spectrometer 2 (CGS2), which contained an array of 7 InSb detectors, thus operating in the same wavelength region as CGS4. The advantages which a 2D array instrument has over the previous generation is easily understood. The larger numbers of pixels available in both the dispersion direction (62) and the cross–dispersion direction (58) allow for increased wavelength coverage and the introduction of a $92\widehat{''}$ slit for spectroscopy of extended areas. This, combined with the increase in sensitivity from the increased wavelength and spatial coverage, gives an overall reduction in exposure time of a factor of about 1600, for background limited observations. In addition to the simple increase in the number of pixels available, detector technology has moved forward resulting in a reduction in

pixel size from the $\sim 200\mu$m pixels used in CGS2 to the 76μm pixels of CGS4. This results in a smaller intrinsic capacitance for the array and a reduction in the effective read–noise. The read–noise is also reduced by improved, low–noise electronics. Both of these greatly enhance the performance of the InSb detector arrays in a high–resolution, read–noise limited regimes.

1.2.1 Long Slit Spectroscopy of Compact and Extended Sources

When installing a 2D pixel array in a spectrometer, the question of whether of not to include a cross–dispersing element must be addressed. An example of a NIR spectrometer which uses this optical technique is the Infra Red Imaging Spectrometer (IRIS) on the Anglo Australian Telescope (Gillingham & Lankshear 1990). IRIS uses a 128×128 pixel mercury–cadmium–telluride array, and has two spectral bands, one in which the I and J windows are observed simultaneously, the second in which the H and K windows are observed simultaneously. The InSb array presently installed in CGS4 has 58 pixels in the spatial direction. The decision against a cross–disperser was based on the fact that the number of pixels was too few to accommodate the different spectra along the array and still give moderate spatial coverage. In the absence of a cross-dispersing element the slit extent is $30\times3.1\widehat{''}$ with the 150mm camera, a total of $92\widehat{''}$ on the sky. This is nicely suited to the scale of the molecular clouds often studied in the NIR, *e.g.* the Orion bar region is $\sim 100\widehat{''}$ long, so the long slit permits moderately high spatial resolution measurements of extended regions. Replacing the current camera with a 300mm focal length camera, reduces the pixel scale to $1.5\widehat{''}$ over $90\widehat{''}$ for still higher spatial resolution.

The provision of a long–slit affects observations of point–like and extended sources quite differently. The most significant improvement for the observation of extended sources is the ability to measure $92\widehat{''}$ (30 spectra) of the source with one exposure. For background noise limited observations with CGS4, the reduction in exposure time required to map an extended source is a factor of 29 compared with CGS2, ignoring the difference in pixel scale ($5.5\widehat{''}$ for CGS2, $3.3\widehat{''}$ for CGS4). Alternatively, sources 5 times fainter can be observed with the same exposure time. More subtle improvements include the facility for determining line strengths or ratios simultaneously at many positions with

exact spatial registration which is invaluable for observing changing conditions across an ionisation front, for example. The observations of the Galactic centre (Chapter 6) make use of this. For the detection of faint lines, a gain in signal/noise of $\sqrt{30}$ can be made by coadding all the rows illuminated by the source. The only limitation to this is the possible error introduced by the gain variations between pixels. This is discussed in the work on flat–fielding the array, Section 2.11.

The long slit also has advantages for observations of point–like or marginally extended sources. The offset beam position for removing sky fluxes, either in nodding or chopping observation modes, can be chosen as a position along the slit. The resulting object–sky pair will resemble that shown in Figure 1.3. For observations in the thermal region of the infrared, longward of 3μm , where sky chopping must be used with three beams defined, this technique can be extended so that the source is measured at three positions on the array. This has three advantages:

1. the sky and object are observed simultaneously, so the problem of temporal variation of the OH line emission is reduced, assuming that the spatial variation is on scale much larger than the slit (Chapter 4).

2. because the object is always observed, and a factor of $\sqrt{2}$ is gained in signal/noise over observations with a linear array.

3. the sky signal remaining after subtraction of the sky frame can be removed from the object by extrapolating from the areas of the slit which are illuminated by the sky alone. There are sufficient "sky rows" for this process even when nodding along the slit.

The predicted sensitivity of CGS4 for observations of a point–like source, nodding or chopping along the slit, is shown in Figure 1.4.

1.2.2 Increased Wavelength Coverage

The choice of resolution for an instrument is driven by the science that is targeted, but also by the grating sizes available and the degree to which the size of cryostat

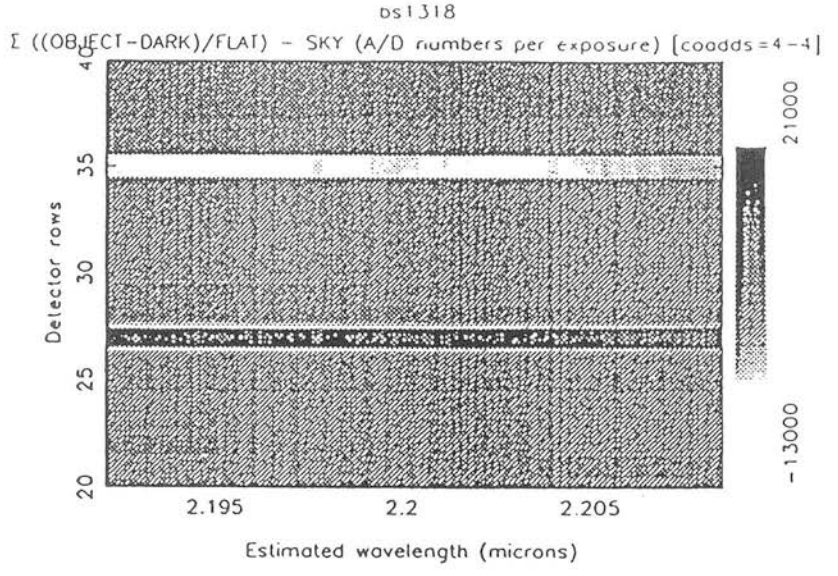


Figure 1.3: Observation showing a standard star, BS1318, which has been observed at two positions along the slit.

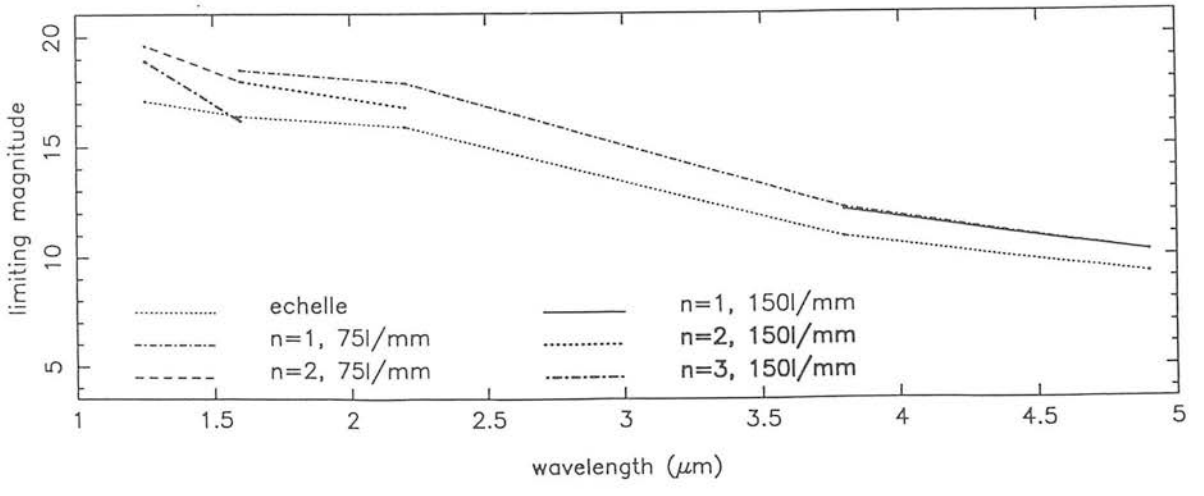


Figure 1.4: The predicted sensitivities for CGS4 on UKIRT for a compact source observed at two positions on the slit. It is assumed that the all the light from the source falls on one pixel.

is constrained by the grating size. In the near infrared, scientific projects could be considered to fall into three resolution regimes:

(*i*) $\frac{\lambda}{\delta\lambda}$ ~400. This covers projects requiring broad band spectra of a number of emission lines, *e.g.* observations of H_2 from photo–dissociation regions, (see Chapter 5), studies of stellar spectra, studies of broad dust features.

(*ii*) $\frac{\lambda}{\delta\lambda}$ ~1000. This moderate resolution provides a capability for studies of, for example, broad lines from galaxies or for separation of H_2 lines.

(*iii*) $\frac{\lambda}{\delta\lambda}$ ~10,000. This mode covers studies of gas dynamics (see Chapter 6) or studies of spectral regions confused by many emission lines or where weak lines are hidden in the wings of stronger lines.

With CGS4 it was decided to provide for each of these categories by supplying three gratings, any two of which may be mounted back–back within the cryostat and can exchanged easily during a night's observing. Of the three gratings are available, two low–dispersion gratings of 75l/mm (R=$\frac{\lambda}{\delta\lambda}$ ~ 200 to 500) and 150l/mm (R=~400 to 1000) and a high–dispersion grating (R~8,000 or a velocity dispersion of 38km s^{-1} per pixel) with a 63.5° blaze angle, of which any two are in the cryostat at a given time due to space constraints. The resolutions quoted are for the gratings with the short focal length camera (focal length = 150mm); twice the resolution is provided by the long focal length camera (focal length= 300mm).

The increased wavelength coverage has a significant impact on the sensitivity of the instrument. For the same resolving power, the gain in wavelength coverage is then a factor of 62÷7=9 over the wavelength coverage of CGS2; for example, the entire K–band from 2.0–2.4μm may be observed simultaneously with the 75l/mm grating with a resolving power, R=350. This same observation would require at least 10 grating positions with CGS2, implying a factor of ten reduction in exposure time for background limited observations with CGS4, compared with CGS2, for this reason alone. However, it is not merely the reduction in observing time which improves the spectroscopy. Simultaneous observations of a range of lines at low dispersion is important for accurately determining line ratios, as temporal changes in atmospheric transparency, and even changes in the slit

position on the sky as the grating angle is changed, can introduce errors. This spectral multiplex advantage is equally applicable to point–like and extended objects.

The larger array allows a high resolution echelle grating to be included, which would have been impractical with smaller arrays. Emission lines from the broad–line region of active galactic nuclei, or from the winds of Wolf Rayet stars are known to be thousands of km s^{-1} broad; H_2 lines from the Galactic centre had previously been measured to be greater than 100km s^{-1} broad (Gatley *et al.* 1986) and the line widths from the ionised gas near the centre are much greater than this. To measure the broadest lines the highest resolution feasible with the CGS2 array would be ~100km s^{-1} , would not permit velocity resolved mapping of the Galactic centre in H_2 . Using the CGS4 array, the echelle resolution of 37km s^{-1} will allow lines of up to ~1000km s^{-1} to be observed without changing the grating position.

1.3 The H_2 Molecule and Fluorescent Emission

Chapters 5 and 6 contain the results from two scientific programmes of observations carried out with CGS4 on UKIRT. In both cases the observations are of the H_2 molecule, and the aim has been to determine the excitation mechanism.

1.3.1 The H_2 Molecule

The H_2 molecule is a homonuclear, diatomic molecule which possesses no permanent electric or magnetic moment. Transitions between the rotational–vibrational energy levels, shown in Figure 1.5 for the lowest vibrational states, are then quadrupole in nature and subject to the selection rule ΔJ=0,$\pm$2: there is no restriction on Δv. Typical energy differences between the states are ~10,000K, corresponding to a wavelength of ~1.5μm . The transition labelling convention usually adopted is

	O(J)	P(J)	Q(J)	R(J)	S(J)
$\Delta J = J' - J$	-2	-1	0	1	2

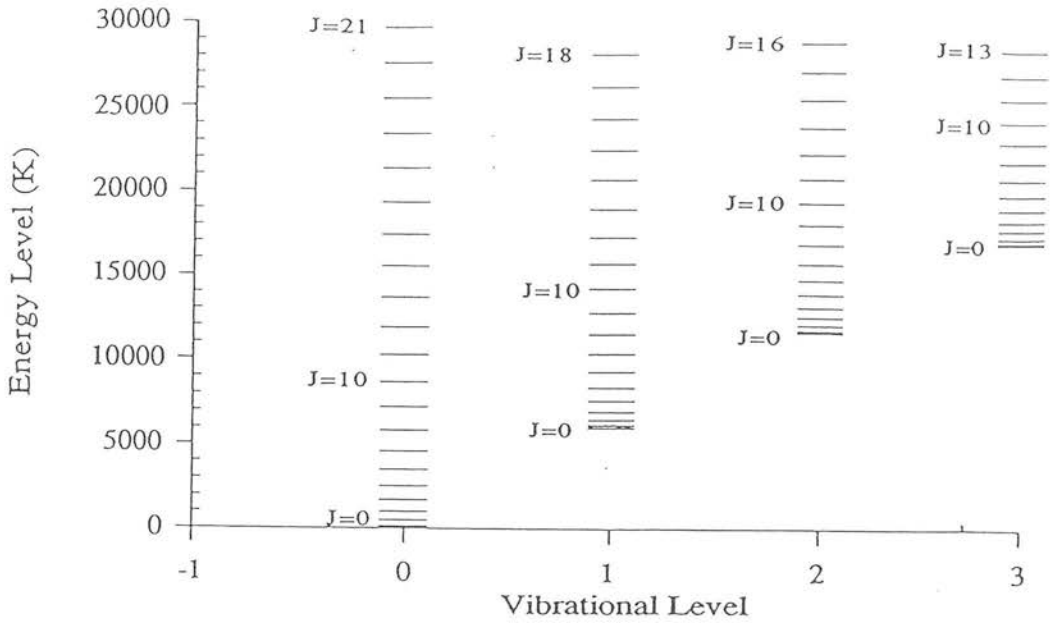


Figure 1.5: The energies of the first vibrational levels of the ground electronic state of the hydrogen molecule.

Thus the allowed transitions for H_2 are the O(J), S(J) and Q(J) transitions.

Each rotational level is further split due to the nuclear spin angular momentum of the molecule, I. The H_2 nucleus has I=$\frac{1}{2}$, giving possible values for the total spin angular momentum of the molecule, T=0,1 and level degeneracies of 3 for states with aligned spins (asymmetric, ortho or odd states) or 1 for states with non–aligned spins (symmetric, para or even states); this ratio (3:1) is referred to as the ortho–para ratio. Figure 1.6 shows the distribution of asymmetric and symmetric states for H_2 .

The total level degeneracy, g_j is given by 3(2J+1) for the odd states and (2J+1) for the even states. The line intensity for emission from any one of these levels is given by

$$I_j = N_j A_j \frac{hc\tilde{\nu}}{4\pi} \tag{1.2}$$

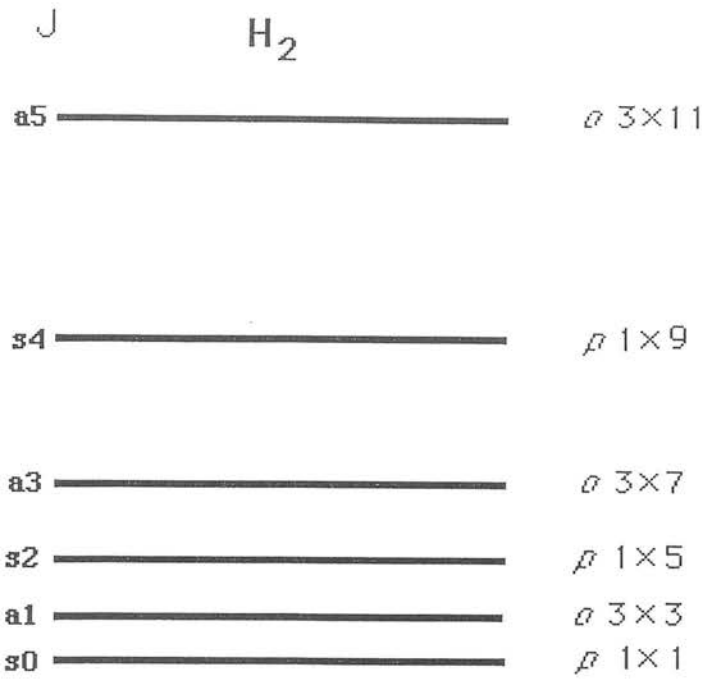


Figure 1.6: The distribution of symmetric and anti–symmetric states for H_2 .

where $\tilde{v}$ is the wavenumber of the line in cm^{-1} (Black & van Dishoeck 1987), A_j is the transition probability (Turner, Kirby–Docken & Dalgarno 1977) and where N_j is the column density of gas in level j, given by

$$\frac{N_j}{N_{tot}} = \frac{g_j e^{\frac{-E_j}{kT}}}{Z} \tag{1.3}$$

where E_j is the energy of the level, j, k is Planck's constant, T the excitation temperature, N_{tot} is the total column density for all states and Z is the partition function, $\Sigma_j e^{\frac{-E_j}{kT}}$. Due to the quadrupole nature of the transitions, the emission lines are inherently weak; the Einstein probability of a transition is A$\sim 10^{-7} s^{-1}$.

1.3.2 Fluorescent Emission from Photodissociation Regions

Several detailed models have been constructed which predict the line intensities of H_2 for given physical and chemical conditions. This thesis is concerned with the models

which pertain to emission from photodissociation regions (PDRs). A PDR is formed when UV light from young, hot stars illuminates the surrounding molecular material from which they have recently formed; it is a region in which the heating and chemistry are dominated by FUV photons (Tielens & Hollenbach 1985). Such a situation is not uncommon and PDRs are found in, for example, planetary nebulae and star–burst galaxies as well as in molecular clouds. When a photon of energy in the range 6eV–13.6eV is absorbed by H_2 , the molecule is excited into the first electronic bands (the Lyman and Werner bands); thereafter ~10% of the molecules are dissociated by de–excitation into the vibrational continuum of the ground electronic state (the photodissociation that gives these regions their name) and the other 90% undergo a transition into one of the bound rotational–vibrational levels of the ground electronic state where they are deexcited by transitions through the rotational–vibrational levels, emitting photons in the NIR.

The fluorescent emission spectrum has been modelled by Tielens & Hollenbach (1985), Black & van Dischoeck (1987) and Sternberg (1988) for low density gas illuminated by moderate UV fields. It is customary to express the UV field strength as a multiple of the ambient interstellar field determined by Habing (1968), G_o=1.6×10^{-3} ergs s^{-1}cm^{-2} (Tielens & Hollenbach 1985). All of these models treat the PDR as a plane, parallel slab of material, static and in equilibrium; the H_2 abundance is assumed to be in steady state with the rate of formation of H_2 on grains being exactly balanced by their destruction. The chemistry in PDRs, and the emission from them, are dependent only on the density of the clouds and the UV field to which they are exposed. Tielens & Hollenbach (1985) have modelled the detailed chemical balance of the majority constituents of a molecular cloud (carbon, hydrogen, oxygen, their isotopes and compounds) and predict the heating and cooling profiles with depth into the cloud and the temperature structure of the cloud for moderate densities and UV fields found in PDRs ($I_{uv} = 10^3 - 10^6$cm^{-3} and $n_o = 10^3 - 10^6$cm^{-3}). The chemical structure of the PDR, showing the transitions between the atomic and molecular species is shown in Figure 1.7. This structure is determined by the competition between the molecules and dust grains for the UV photons, and is sensitive to the degree of absorption by CO, absorption by H_2 (H_2 /CO self–shielding) and the physics of the dust scattering (dust–shielding). The scat-

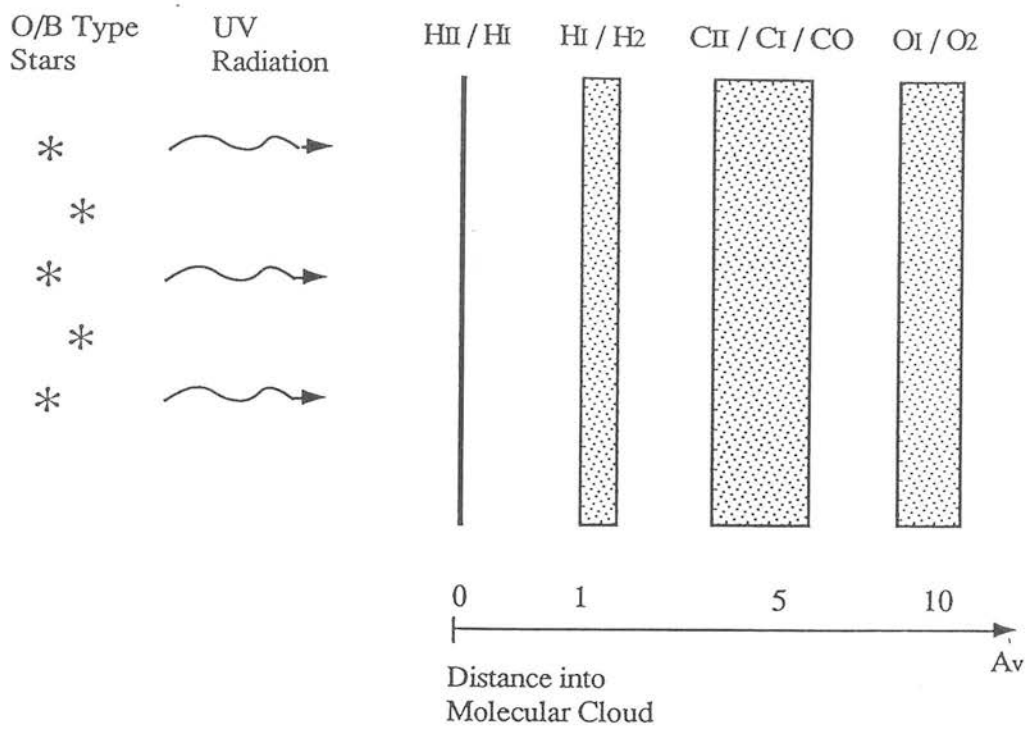


Figure 1.7: A cross section of the structure of a PDR, showing the depths into the cloud at which the transitions between atomic and molecular species occur (from Tielens & Hollenbach 1985).

tering function for the dust is a major uncertainty; it has the form $I \propto e^{-k_\lambda A_\lambda}$ with k taken as 1.35 (Tielens & Hollenbach 1985) but great uncertainty exists in this value. Tielens & Hollenbach (1985) modelled the PDR for a greatly simplified H_2 molecule by considering only the molecule in its ground state and or excited into a pseudo–level (v=6) by UV photons or by formation on grains. A more complete treatment of the details of the fluorescent cascade has been carried out by Black & van Dishoeck (1987) and Sternberg (1988) for low density conditions with moderate UV fields, modelling the radiative transfer of the UV pumping into each of the the 14 bound vibrational levels of the ground state. The model spectrum derived for Black & van Dishoeck's basic model, Model 14 (n=3×10^{-3}cm^{-3}, I_{uv}=10^3), is shown in Figure 1.8. Since the levels are populated by transitions from the excited electronic states and by radiative de-excitation of higher levels in the ground electronic state, the column density of gas in the high

rotational–vibrational levels with E_j ~20000K are far greater than they would be for thermally populated levels. Although the total emission intensity is dependent on the model parameters, the relative intensities of the lines depend on the details of the radiative transfer, and are insensitive to the density and UV field over the parameter space modelled by Black & van Dishoeck. For comparison, Figure 1.8 also shows a thermal emission spectrum from a 2000K gas excited, typical of the spectrum which would be observed from H_2 excited in collisions due to the passage of a shock–front. The lower levels such as 1–0 S(1) (6952K) and 2–1 S(1) (12552) show strong emission lines, whereas the higher levels are not populated and show no significant emission. Thus the 1–0 S(1)/2–1 S(1) ratio, which is 1.7 for fluorescent emission and ~10 for thermal excitation, came to be adopted as a straightforward diagnostic of the excitation mechanism.

The extension of these theories for higher density regions and those illuminated by higher UV fields, effectively invalidates the use of the 1–0 S(1)/2–1 S(1) ratio as the sole diagnostic of the excitation mechanism (Sternberg & Dalgarno 1989; Burton, Hollenbach & Tielens 1990). Burton, Hollenbach & Tielens have recently expanded this work to include densities in the range 10^5–$10^7 cm^{-3}$ and UV fields ~ 10^5–10^7 times the interstellar field. They show that for $n > n_{crit}$, defined

$$n_{crit} = \frac{A}{\gamma} \tag{1.4}$$

where A is the rate of radiative decay and γ is the rate of collisional de-excitation, the effect of collisional de-excitation on the fluorescent cascade becomes significant. As the radiative transitions are preempted by collisional de–excitation, the spectrum ceases to be characteristic of UV excitation, and the population of the lower levels is increased. This collisional de–excitation also contributes to heating the gas, giving some thermal contribution to the lower level populations. A plot of column density against energy level for both a shock–excited gas and a radiatively excited gas shows clearly the different ways in which the energy levels are populated for radiatively excited or thermally excited gas (Figure 1.9). Note that in the fluorescent case, for the high v levels, the points lie on distinct curves, whereas for the shocked gas all the points lie on a curve determined by the gas temperature. The lower v levels for the high–density gas have become "thermalised";

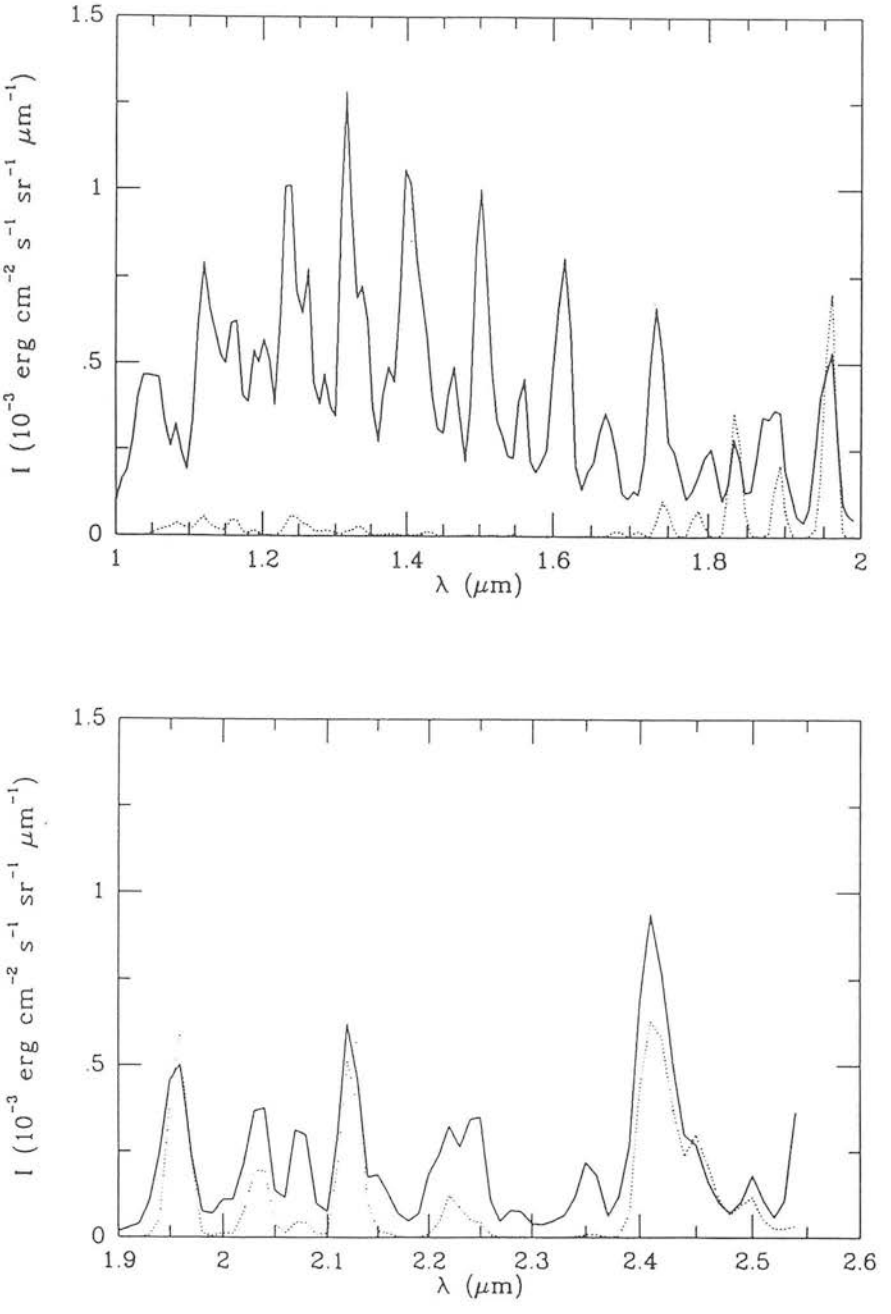


Figure 1.8: The theoretical spectrum for the emission from H_2 , as per Black & van Dishoeck's Model 14 (solid line) and for a thermal emission spectrum for a gas at 2000K (dashed line). Note that the strengths of the higher vibrational level H_2 lines are greatly reduced, and specifically that the strength of the 2–1 S(1) is far greater in the fluorescent spectrum.

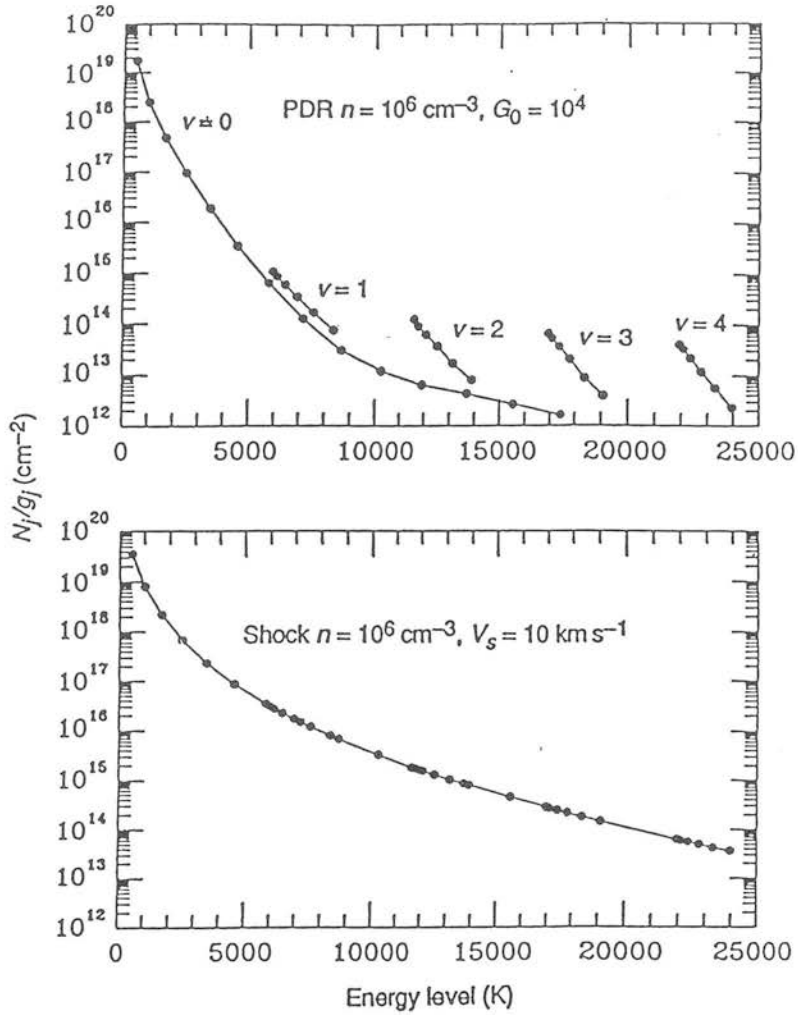


Figure 1.9: Column density of levels against energy for radiatively and thermally excited gas (Burton 1992)

they lie on the line defined for the shocked gas. Thus the 1–0 S(1) lines and 2–1 S(1) are susceptible to thermalisation in a high–density gas, and may give misleading predictions for the excitation mechanism. The emission from the higher v levels maintains the characteristics of a fluorescent spectrum, and is a more reliable indicator.

Burton, Hollenbach & Tielens also show that the effects of increasing density and UV field on the structure of the PDR. For a model where the attenuation of FUV photons by absorption by dust grains dominates, the structure shown in Figure 1.7 is assumed. However, by assuming that the effect of dust is negligible compared with the H_2 self–shielding and that the formation and destruction of H_2 are balanced, it is possible

to derive the conditions for line absorption of UV photons by H_2 (self–shielding) to determine the PDR structure. This will occur when

$$\frac{n}{G_o} \geq 40\text{cm}^{-3} \tag{1.5}$$

For this regime, HI/H_2 and CII/CO transitions (Figure 1.7) will be brought closer to the PDR surface, exposing the dense gas to the full heating effect of the UV field and increasing the population of the lower levels.

Chapter 5 contains an study of the H_2 emission from the planetary nebula Hubble 12, which was identified as a source of fluorescent emission by Dinerstein *et al.* (1987). Early observations for the Galactic centre (Gatley *et al.* 1984, 1986) gave a line ratio for the 1–0 S(1)/2–1 S(1) of ~10 which was interpreted as arising from shock excited gas. In the light of the theories outlined above, we undertook to repeat this experiment with the aim of confirming the existence of fluorescent emission which is implied by the strong emission from other species associated with PDRs (*i.e.* [CII] 158μm and [OI] 63μm). The determination of the excitation mechanism will have an impact on the understanding of the central source in the Galaxy: evidence for both a burst of star formation and accretion or ejection of matter from a massive point–like central source is to be found in the results from the intensive investigation of this object reported in the literature. The results from observations with CGS4 are presented in Chapter 6.

Chapter 2

Cooled Grating Spectrometer 4

Preface

The work summarised in this Chapter was carried out by the members of the CGS4 project team over many years, and is included to give a full account of the design and operation of the instrument. The author's contribution to this has been the work on the sensitivity of the instrument (Section 2.8), the investigation of flat–fielding methods and results thereof (Sections 2.11– 2.12).

Summary

As one of the first of a new generation of infrared instruments, Cooled Grating Spectrometer 4 capitalises on the recent availability of NIR–sensitive 2D arrays, and the design was driven by the need to maximise the astronomical gains. The requirements of the scientific projects which would benefit from the new instrument included: high throughput, reliable calibration, good image quality, flexibility, comprehensive wavelength coverage and a long–slit capability allowing good coverage of extended sources or simultaneous sky monitoring for point–like sources. The instrument should be cooled to give maximum sensitivity for the NIR (Section 2.4). This Chapter is an introduction to CGS4; aspects of the design and the way in which they apply to the goals listed above are discussed. The

requirements for producing a flat–field and the way in which they are met are discussed. Details of other measurements required for calibration, such as the measurement of the dewar background are included.

2.1 The Optical Design

The layout of the CGS4 optics is shown in Figure 2.1. The optical path can be described as follows. An f/36 beam from the telescope comes to a focus inside the cryostat. The beam is then collimated by an arrangement of three mirrors which also act as an image rotator, allowing any slit orientation on the sky. The beam passes through a Lyot stop in the pupil plane, which acts as a cold baffle. After passing through the filter wheels the beam is converted to f/10 by a BaF_2/CaF_2 quasi–achromatic doublet and brought to a focus at the slit wheel which contains a range of slits to maintain the alignment of the slit image with a column on the detector array (see Section 2.3). The filter wheels contain brad–band filters covering 1.3–5μm , a set of narrow band filters for wavelengths and a circular variable filter (CVF), with a 2% passband for order–blocking the echelle. The beam is re–collimated before being dispersed by one of two gratings mounted back–back, then finally brought to a focus in the plane of the array by one of two interchangeable cameras of different focal length.

In this Section, the way in which the details of the optical design were determined is discussed. The basic design is that of a Czerny–Turner spectrometer, but with a quasi–Littrow configuration; the incident beam strikes the grating at an angle of 14° out of the plane of dispersion. A Littrow configuration has the advantage of making the optics more compact, *i.e.* reducing the size of the cryostat which is to be cooled, and adopting an off-axis configuration reduces the effects of scattered light which would be seen with a classic Littrow mount. Modelling of the optical performance (Section 2.2) demonstrated that the optical design used for CGS4 could provide sufficient image quality to allow the use of this configuration. One disadvantage of the quasi–Littrow mount is that the changing out–of–plane angle along the slit causes the off–axis rays to suffer a distortion: images of the slit appear to be sheared on the array and the footprint of the beam diffracted by the echelle to be elliptical in cross-section. The origin of this effect is described in

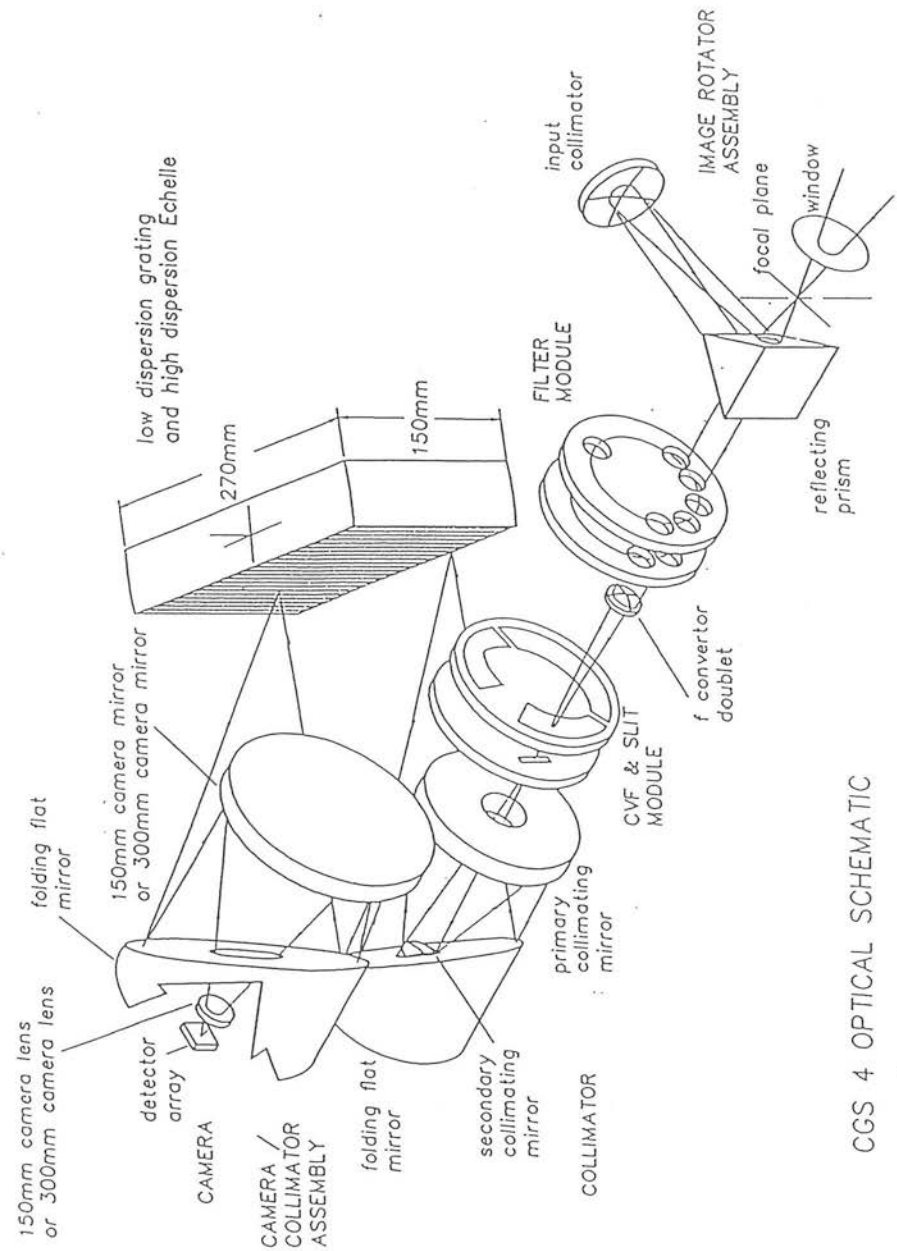


Figure 2.1: A schematic diagram of the optical layout of CGS4, showing the optical components and the modular nature of the design.

Section 2.3.

When determining the details of the optical design, the initial parameter to be set is the pixel field of view, and a balance must be achieved between the desire for high spatial resolution/small pixels, the speed of the optics and the limitations imposed by the seeing and autoguiding. The seeing disk for UKIRT on Mauna Kea is typically 1 $\widehat{''}$, rising to 2 $\widehat{''}$ in poorer weather. The beam size is chosen to be approximately the size of the seeing disk; a compromise must be reached between achieving high signal/noise using a small enough slit to reject most of the background, and retaining photometric accuracy by using a slit which is large enough to gather most of the light from an unresolved source. Since the design for CGS4 incorporates two cameras, the design of the long focal length camera was set to give a pixel size of $1.5\widehat{''}$ on the sky, the smallest value deemed practical, and the short focal length camera designed to give $3\widehat{''}$ pixels. These are the fields of view chosen for the 76μm pixels. The design of CGS4 was also influenced by the desire to accommodate 256^2 pixel arrays of 30μm pixels, giving a field of view of $1.5\widehat{''}$ with the short focal length camera and $0.6\widehat{''}$ with the long focal length camera.

The f–numbers for the cameras are determined by the need to conserve the étendue of a system. Etendue, or throughput, is given by

$$G = A\Omega = 2\pi A\,(1 - \cos(\Phi)) \tag{2.1}$$

where A is collecting area of the beam in a focal plane, Ω is solid angle subtended by it and Φ is the half–angle of the cone of light.

For any ideal optical train, this quantity must be conserved to maximise the efficiency of the instrument. For a given FOV in arcseconds, telescope diameter D, and pixel size, w, the f/# is completely defined by the conservation of étendue:

$$D\frac{FOV}{206265} = \frac{w}{f/\#} \tag{2.2}$$

For the optical system defined by UKIRT and CGS4, with pixel width, w=76μm , the telescope diameter, D=3.8m and FOV of $3.1\widehat{''}$ the camera f/# is f/1.35; for a 1.5 $\widehat{''}$

FOV is f/2.7. The spectral resolution required (Section 1.2.2) and the choice of gratings available commercially then define the camera focal length. The spectral resolving power of a spectrometer is given by:

$$R = \frac{2f}{w}(\tan\theta_b) \tag{2.3}$$

where f is the camera focal length, and θ_b is the grating blaze angle. In practice the focal length was determined by the echelle grating. The most readily available blaze angle for echelle gratings is 63.5°, so for a resolving power of ~10,000, the camera focal length must be ~0.2m, giving a beam size on the camera of ~0.15m. This is also the size of the footprint on the grating, so one dimension of the grating is fixed. The second dimension is also fixed at twice this size, as it must accommodate the circular beam at a grating angle of 63.5°. For a given telescope diameter, the echelle grating and the choice of beam width in arcseconds effectively define the entire optical design and the size of cryostat required. The 75l/mm and 150l/mm low–dispersion gratings were chosen from those available commercially which would provide the resolving powers required for the astronomical projects of interest. A graph showing that the spectral resolution required can be provided by the chosen gratings is given in Figure 2.2 (Mountain *et al.* 1990). For the 1.5″ pixel, f/2.7 configuration the camera focal length is 300mm. The other component to be specified is the f/convertor lens which focuses the collimated beam which has passed through the filter wheel at the slit plane. After collimation by the image rotator, the f/36 beam from the telescope is narrow enough to be accommodated by the 1–inch filters, but the f/convertor should expand the beam to fill the grating. The required f–number for this lens to work with the components already defined is f/10. A diagram of the optical design showing the way in which the étendue is conserved is shown in Figure 2.3. The width of the slit is determined by the slit position, by both the camera and collimator focal lengths, by the size of the pixels and the number of pixels that the slit is matched to. With the CGS4, the slit width (0.56mm for the low–dispersion grating, 0.45mm for the echelle) is exactly matched to one pixel (76μm), the optimum number of pixels per slit width from signal/noise considerations. This introduces the need to oversample the spectra, by moving the detector array relative to the slit to fully define the spectral line shape (Section 3.11).

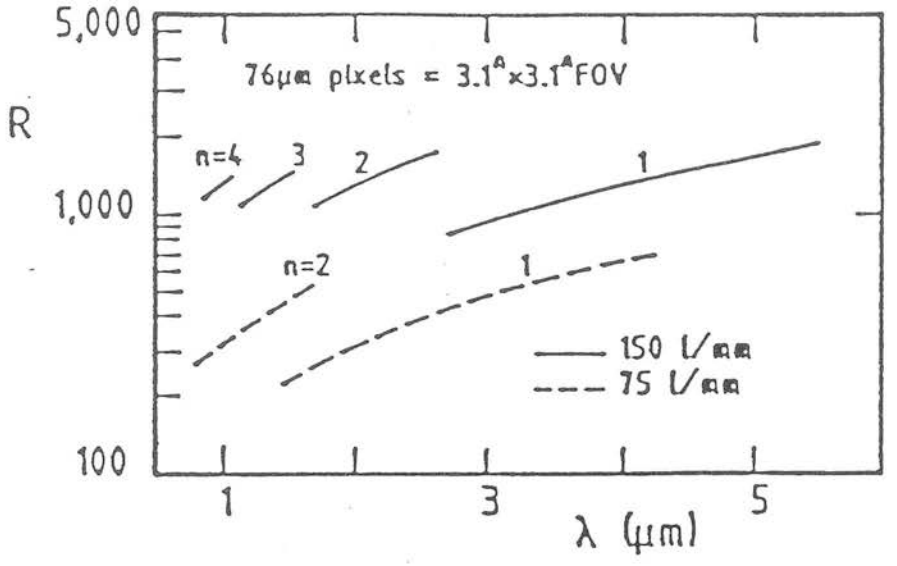


Figure 2.2: Resolving power as a function of wavelength with both the long and short focal length cameras for the 150l/mm and 75l/mm gratings chosen for CGS4.

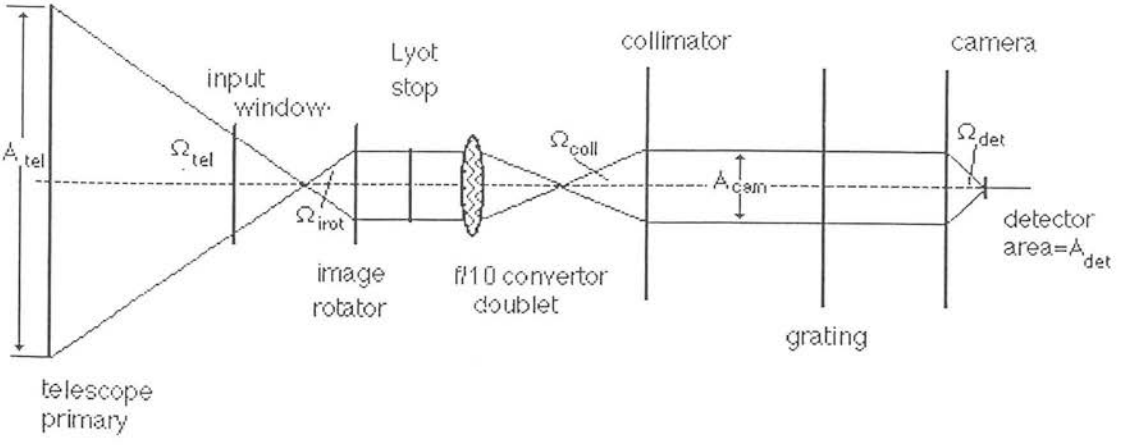


Figure 2.3: The optical layout of CGS4, showing the principle of conservation of étendue along the optical path.

	Power (Fringes)	Irregularity (Fringes)	Clear Aperture* (mm)	Tilt: (mrad)	DEC: (mm)
1st flat	1	1/2	91.8	0.27	0.1
Input Collimator	1	3/4	69.4	-	0.05
2nd flat	1	1/2	34.3	0.27	0.1
F/converter lens	4	1	30	1	0.1
Secondary Collimator	2	3/4	28.4	0.49	0.045
Primary Collimator	2	3/4	134.1	0.27	0.025
1st compound flat	1	1/2	142.6	0.34	0.025
Grating	3	1/2	143x258.7	0.72	0.1
2nd compound flat	1	1/2	~250	0.34	0.025
Camera Mirror	4	1/2	~250	0.27	0.035
Camera lens:					
1st surface	1	1/4	39.2	0.41	0.053
2nd surface	1	1/2	39.2		

Table 2.1: The tolerances on the positioning of the CGS4 optics.

2.2 Defining the Mechanical Tolerances and Modelling the Optical Performance

Since CGS4 was designed to accommodate both the 58×62, 76μm pixel arrays and the 30μm pixels of the 256^2 arrays, the optical tolerances were set by the acceptable degradation in image quality for the smaller pixels and larger arrays. A spot–size of 30μm (80% encircled energy) at the detector gives a 10% degradation of image quality with the 76μm pixels. This spot size should be maintained images at the edges of the array, *i.e.* ±4.5mm from the optic axis in the detector plane for the smaller arrays and ±9mm for the 256^2 arrays. The tolerances on positioning the optical components to meet this stipulation were calculated by the optical modelling programme, CODE V, and are given in Table 2.1. Expected spot sizes for on and off–axis rays were also calculated using CODE V; an example is shown in Figure 2.4. The tolerances are more stringent than could be achieved by traditional optical alignment and maintained on cooling to 77K, as required to achieve the high sensitivity sought for CGS4. To preserve the geometrical arrangement of the optical components it was necessary to eliminate differential contraction on cooling. This lead to the all aluminium design for the CGS4 optics (Section 2.6).

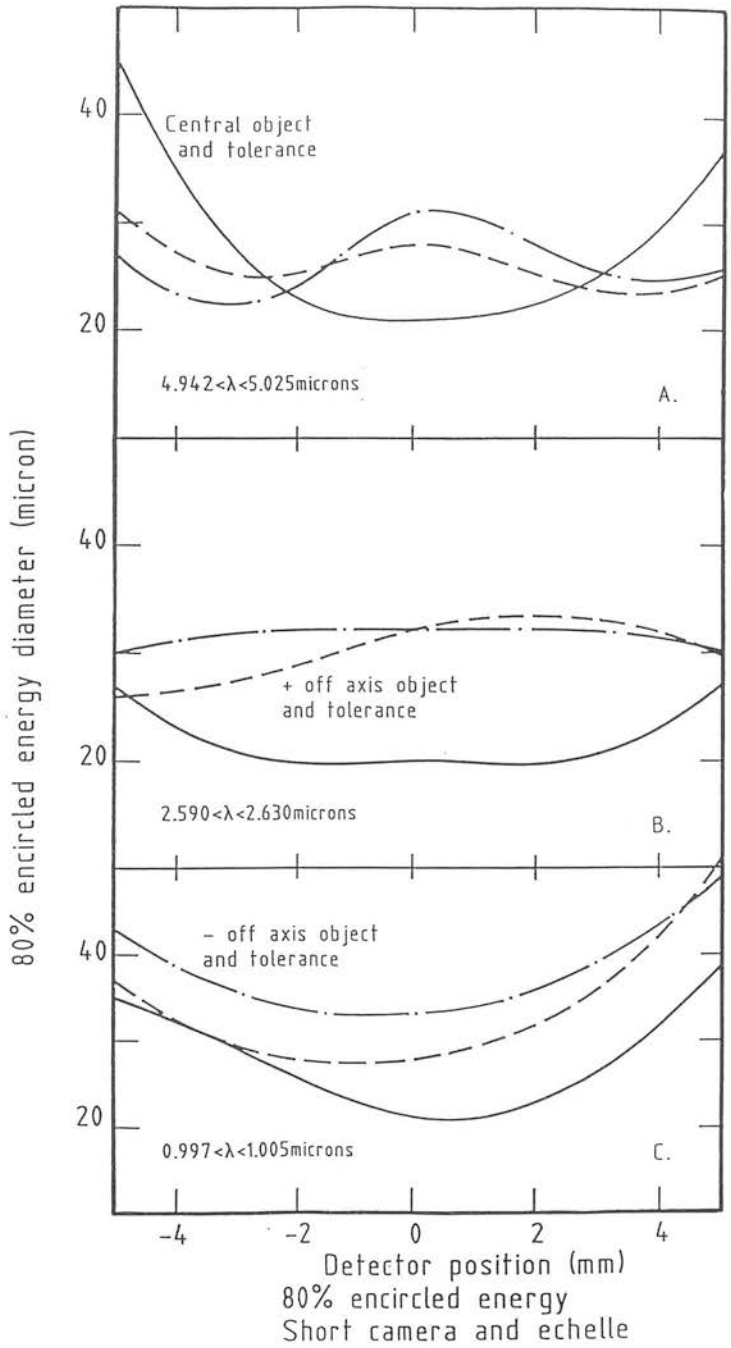


Figure 2.4: The image quality modelled by the CODE V package for on–axis and off–axis beams with the CGS4 echelle and short focal length camera.

2.3 Field Rotation Effects

The off–axis design of the spectrometer causes the image of the slit on the array to be sheared at some angle relative to the columns. The physical cause for this will be seen by considering a ray from the slit as the sum of two vectors, one in the plane of the grating running perpendicular to the rulings and the second orthogonal to this, also in the plane of the grating. The ray which lies parallel to the rulings is reflected according to Snell's law, whereas that running perpendicular is diffracted by an amount given by the grating equation; vector OP is sheared on diffraction at the grating (Figure 2.5). The effect is to rotate the slit through an angle given by

$$\tan(\theta_{slit}) = \lambda \frac{\delta\theta_{grat}}{\delta\lambda} \tan(\gamma) \tag{2.4}$$

where θ_{slit} is defined relative to the cross-dispersion direction. This expression can be derived as follows. A point, P, along the slit has a position $\left(\delta x_p, \delta y_p\right)$ relative to the nominal optic axis O at (0,0), where

$$\delta x_p = \frac{\tan \delta i}{f_{coll}} \qquad\qquad \delta x_p = \frac{\tan \delta\gamma}{f_{cam}} \tag{2.5}$$

and δi is the change in the angle of incidence for a ray from P, $\delta\gamma$ is the change in the off–axis angle for a ray from P and f_{coll} is the focal length of the collimator. At the grating, the vertical ray is reflected with $\delta i = \delta r$, giving a position for P′, the image of P on the detector,

$$y'_p = y_p \frac{f_{cam}}{f_{coll}} \tag{2.6}$$

The horizontal position of P′ will be given by

$$x'_p = \frac{\tan(\delta r)}{f_{cam}} \tag{2.7}$$

The slit rotation angle is defined to be the angle between the vector OP′ on the array and the vertical; it is then given by

$$\tan\theta_{slit} = \frac{f_{coll} \tan \delta r}{y'_p} \tag{2.8}$$

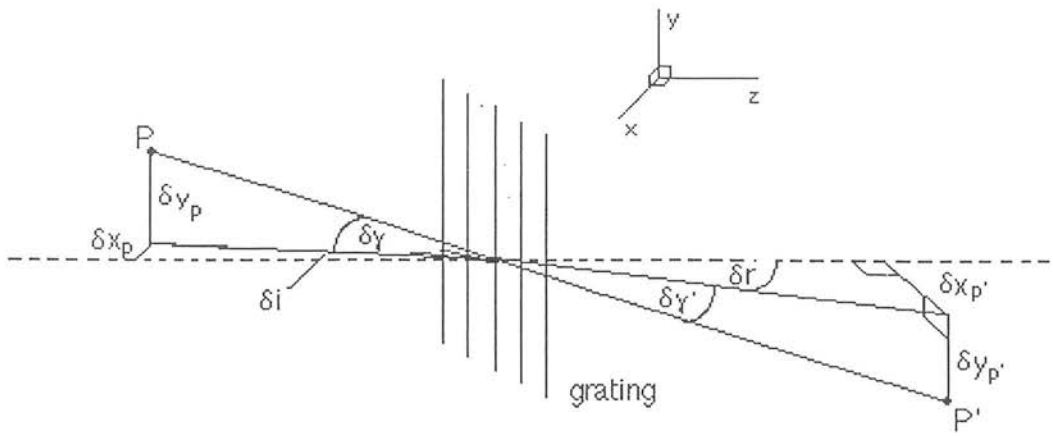


Figure 2.5: Illustration of the rotation of the slit image on diffraction by the grating. The CGS4 grating is drawn as a transmission grating for simplicity.

Replacing y'_p with the expression from Equation 2.6, and taking the limit as δr and $\delta\gamma \rightarrow 0$,

$$\tan(\theta_{slit}) = \frac{\partial r}{\partial \gamma} \tag{2.9}$$

which reduces to the expression given in Equation 2.4. This effect is corrected by the provision of a number of different slits, angled with respect to the optic axis, mounted in a slitwheel which is rotated to give a vertical slit on the array for any given grating angle. A range of slit angles was chosen, to provide for the expected range of grating angles for the two low dispersion gratings and the echelle; these range from -4° to 36.9°, the latter being for the echelle. The slit wheel is discussed further in Section 3.5.

The rotation of the slit wheel causes a corresponding rotation of the slit on the sky; the image rotator can correct for this by introducing a counter–rotation. This any slit–orientation on the sky is possible, and the slit will always be aligned to the detector columns. The slit–rotation also distorts the shape of the pixels on the sky, an effect called anamorphic magnification. The degree of stretching of the pixels is given by $\cos(\theta_{slit}$' it

is negligible for the low–dispersion gratings but gives pixels which are $3.8''\times2.5$arcsec for the echelle with the short focal length camera.

A second order effect of the above is that the projection of the slit onto the array is not only sheared, but the image is no longer a straight line, but a curve. For CGS4, with the short focal length camera and the 58×62 pixel array this effect is negligible. However, with the long focal length camera there is evidence of curvature of the slits on the array (0.2–0.5 pixels). This will be a greater problem for the larger 256^2 arrays.

2.4 The Case for a Cooled Cryostat

The sensitivity of an infrared instrument is improved in two ways if it is cooled: firstly, the thermal background on the detector is reduced if there are no warm, emissive objects in the optical path, secondly the detector performance is enhanced at low temperatures due to the reduction in dark current.

The thermal background on the detector will be photons from a black–body at the dewar temperature. As the emission comes from optical components after the filters along the optical train, the light is unfiltered and the BB emission must be integrated over the 1–5μm region over which the array is sensitive. The detector field of view will be 2π. The dewar temperature required to maintain this thermal background at a level where it does not contribute to the noise on an observation can be estimated by specifying that it should be 10% of the lowest expected sky–noise signal. The lowest sky–background is found for observations with the echelle at 1μm ; the noise from the sky will be ∼10 photons s^{-1} giving a limit on the acceptable noise from the dewar of $1e^{-}s^{-1}$. The temperature implied by this constraint is ∼80K.

Previous IR instruments have been partially cooled, eg the CGS2 cryostat contained some optical components, detector and filters, but much of the optics were warm. These instruments were constrained by the use of liquid nitrogen (LN_2) evaporation to cool them which is expensive in cryogens, gives a short hold–time and limits the size of dewar. With CGS4, the use of mechanical closed cycle coolers gives a much longer cryogen hold–time and allows a larger cryostat to be cooled so that all the optics may be held at LN_2

temperatures. The cooling occurs in two stages, since the optimum array operating temperature is less than 70K. The first stage of the cooler holds the radiation shield, casting and optics at ~70K, while the second stage cools the array housing. The array would reach thermal equilibrium at around 25K, which is cooler than desired so a heater maintains the array temperature at 38K. Since all the mirrors are inside the cryostat, the only thermal background directly observed by the detector is the sky and telescope. Scattered light is minimised by the use of extensive baffling at the input window and in the image rotator unit. Any surfaces which were not reference surfaces for the optical alignment were roughened and pointed with NEXTEL, a black paint, to further reduce scattering.

To reduce the thermal paths into the dewar and light leaks, and hence maintain the low thermal background, the motors required to drive the eight motors which control the positioning of the grating, filters (2), slit, CVF, image rotator and detector (2) were located inside the cryostat. Thus, one of the main technical problems that had to be overcome was how to operate stepper motors at LN_2 temperatures. Any conventional motor lubricant would freeze at 70K, and lubricating the motors is essential to prevent them stalling or loosing steps, with a resulting loss of calibration. The lubricant finally adopted is a graphite compound which lines the bearing race and coats the bearings as they rotate. The tests on reliability and repeatability of the motors carried out during the laboratory commissioning illustrate the success of this approach (Chapter 3). For example, it was found that the grating position was repeatable to 0.1pixel during the equivalent of a years' observing.

The final step in controlling the instrument environment was to encase the entire casting and radiation shield in super–insulation. Many layers of the reflective foil Mylar are used; these are then covered with polyester, which is stitched to the Mylar and attached to the instrument with vacuum tape. The measured cooling curve for the cryostat is shown in Figure 2.6; the total cooling time for the instrument is 24hours.

The proof of the effectiveness of cooling the mirrors and housing the motors in the cryostat is seen in the very low background inside the dewar. Section 2.5 reports that the measured dark current was found to be $\sim 1e^{-}s^{-1}$ (Figure 2.7). The component from

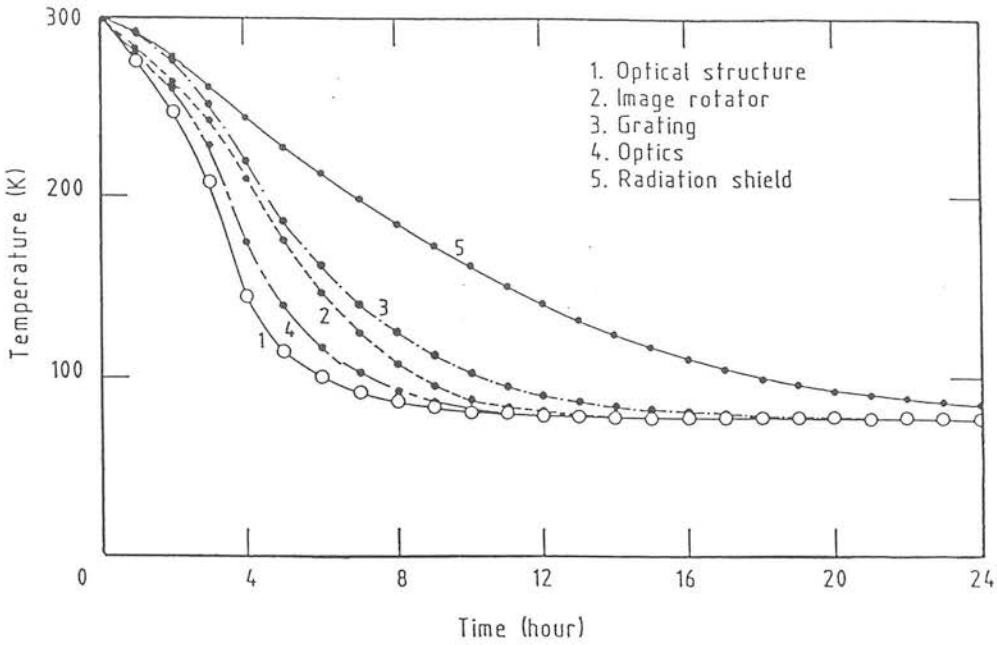


Figure 2.6: The measured cooling curve for CGS4.

the detector and the dewar are inseparable, but this value is entirely consistent with the dark current expected for an InSb detector alone, so we can conclude that the thermal background is negligible.

2.5 The Detector

The CGS4 array is an Indium Antimonide (InSb) device, designed by the Santa Barbara Research Corporation (SBRC) to be optimised for astronomical applications, taking into account the need for high quantum efficiency, low detector and read–out noise and large dynamic range. InSb has a band gap of 0.16eV at room temperature, decreasing to 0.24eV as the detector is cooled, giving an upper wavelength cut–off of $\sim 5\mu$m at the operating temperature used in CGS4. The array architecture is shown in Figure 2.8. It consists of an n–type substrate of donor concentration $\sim 1\text{–}2\times10^{15}\text{cm}^{-3}$; the p–type layer has an impurity concentration of $1\times10^{18}\text{cm}^{-3}$. The electrical connection between the array material and the Silicon read–out layer is made through "bumps" of indium (this is known as "bump–bonding"). Each pixel is read out through an individual source–

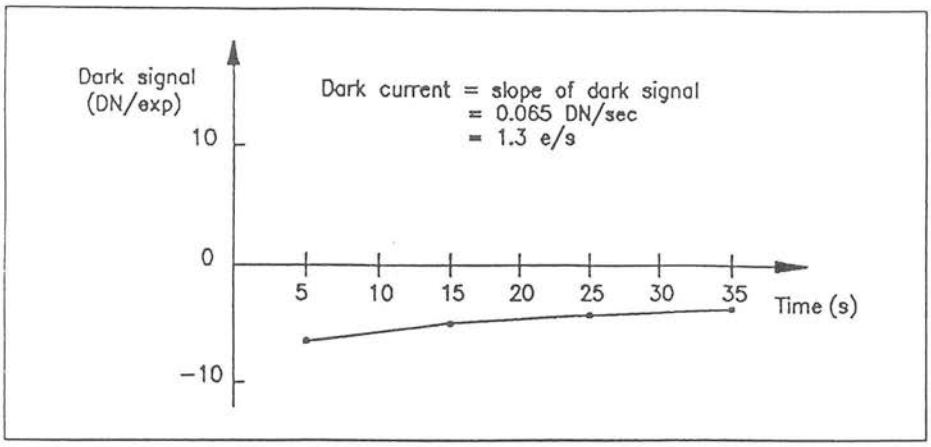


Figure 2.7: The dark current for CGS4, measured using darks in NDR mode, is $\sim 1e^{-}s^{-1}$

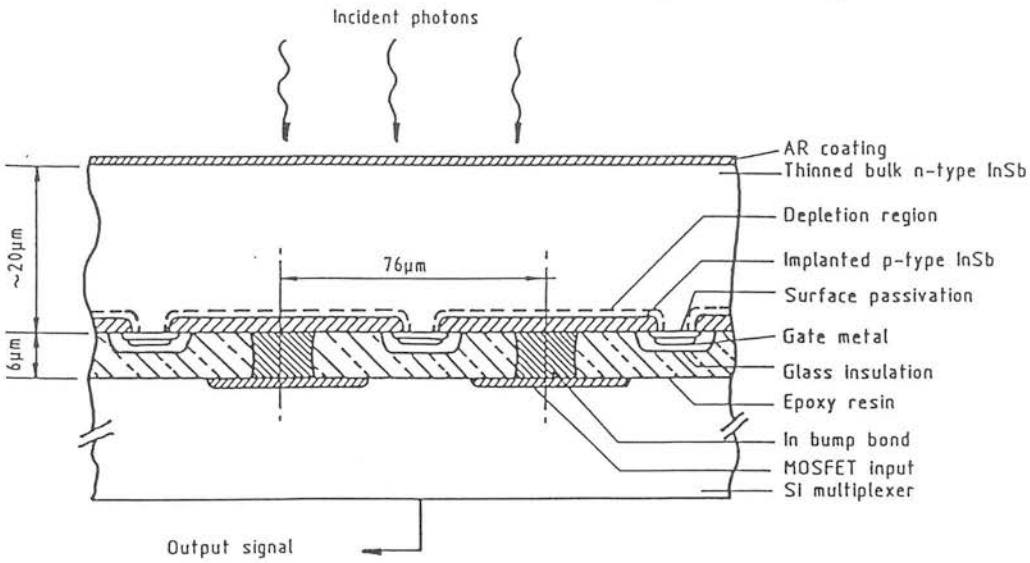


Figure 2.8: A cross-section of the SBRC InSb array shows the general construction, including the n and p–type substrates and the Indium bump–bonds which make the electrical connection between the array and the read–out electronics. This diagram is from from Rayner (1988).

follower circuit. The use of source–follower circuits in the SBRC array has the advantage that the array need not be reset after a read, allowing a variety of noise reducing read–out techniques to be employed. The operation of a source–follower circuit can be explained briefly as follows: the source–follower consists two identical MOSFETS in series which are assumed to have the same characteristics (Figure 2.9). The voltage across the pair is the drain voltage, maintained at 1.5V or less to suppress avalanche emission from the output transistor which has been observed in tests of these arrays (Fowler *et al.* 1987). The gate–source voltage across the lower FET, which determines the drain current through both FETs, is 1V. Hence the voltage across the upper FET is determined by

$$V_o = V_i - 1V \tag{2.10}$$

if no current flows at the output. Thus, the output (*source*) volts *follows* the input volts with an offset determined by the fixed gate–source volts, with no current drawn from the array. In this manner, the array can be read without destroying the charge accumulated on each pixel. The non–destructive read algorithm used with CGS4 is described in Section 2.5.2.

2.5.1 Detector noise and Temperature behaviour

In Section 2.4 it was stated that the optimum operating temperature for the array was 38K. The reason for this will be seen in the following description of inherent detector noise. The main sources of background noise in a low background application such as astronomy are the detector read–noise and the generation–recombination noise, often referred to as dark–current, I_d. The latter is dealt with first.

Dark current includes both electrons generated by the detection of photons emitted by the surroundings which have energies larger than the band gap by virtue of their thermal energy, and generation–recombination (GR) noise. The thermal contribution was discussed in Section 2.4. GR noise results from the recombination of an electron–hole pair before the electron has diffused across the junction and been detected. The

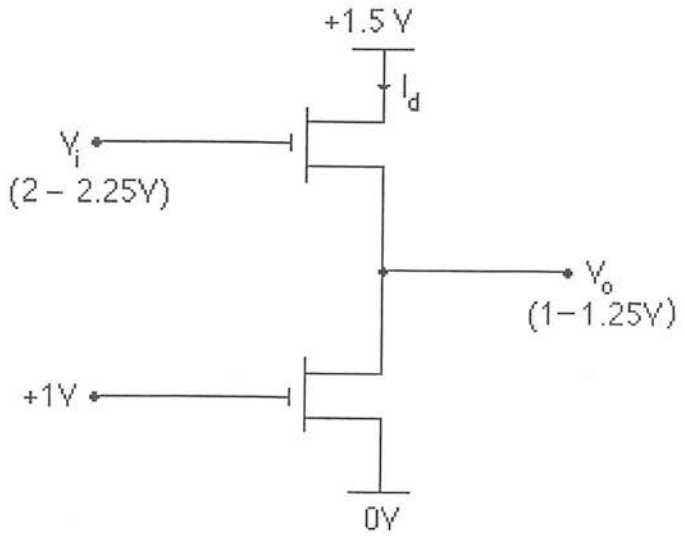


Figure 2.9: A diagram of the source follower circuit used to read out each pixel of the CGS4 array. The input voltage, V_i, can be sampled, without destroying the charge on the array, by measuring the output voltage, V_o.

GR noise can be calculated

$$\mathrm{GR} = \sqrt{\frac{2\mathrm{kTt}}{\mathrm{R}_o}}\frac{1}{e} \tag{2.11}$$

where k is Boltzmann's constant, T is the temperature, t is the exposure time, e is the electronic charge and R_o is the detector resistance at zero bias ($\sim 6\times10^{11}\Omega$). The GR noise is actually dependent on temperature in two ways: firstly through the direct $\mathrm{T}^{\frac{1}{2}}$ dependence and secondly through the fact that R_o increases with decreasing temperature. Thus the sensitivity of the array will increase with decreasing temperature.

The second noise source, read–out noise, is not temperature dependent, but is a function of the read–out electronics. The noise at the output of the array is an rms voltage of 85μV from the multiplexer. For the the capacitance of the 76μm detector pixels $\sim$0.5pF, and the the gain of the source–follower $\sim$0.7, this is equivalent to a charge on the array of $\sim 400e^-$. This is the effective read–out noise of the array.

These are limiting noise sources in observations where the source signal is weak and the background low, *e.g.* observations of faint sources at high spectral resolution. The GR noise can be reduced by cooling the detector; a method for reducing the effective read–noise is described in Section 2.5.2. A lower limit on the detector temperature is given by the rapid decrease in quantum efficiency (QE) with decreasing temperature. The final operating temperature of 38K is a compromise between the dark current and QE considerations.

2.5.2 A Noise Reduction Algorithm

The integrating detector voltage is directly proportional to the incident flux for the linear regime (80% well) in which most observations are made. Figure 2.10 shows the change in flux on the array during an integration.

"Sampling" this voltage, by reading the array without destroying the accumulated charge, and fitting a straight line through the samples using a least squares fitting algorithm results in a reduction of the read–out noise. The slope of the line in Figure 2.10 is found

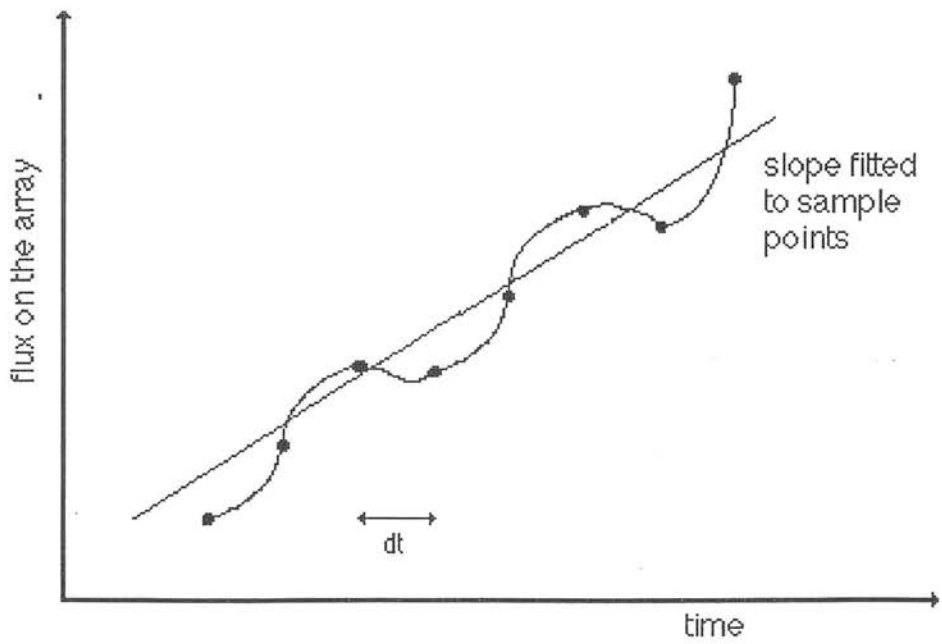


Figure 2.10: The flux on the array during an integration. By fitting a straight line through many sample points, an estimate of the flux on the array is obtained which is more accurate than that for a single read and hence the read–noise is reduced.

from:

$$slope = \frac{\sum_{i=1}^{n} V_i \left[i - \frac{(n+1)}{2} \right]}{dt \frac{n}{12} (n^2 - 1)} \tag{2.12}$$

(Chapman *et al.* 1990) where V_i is the voltage of the ith sample, n is the total number of samples and dt is the time between samples; it is assumed that dt is fixed and known in advance. The reduction in the error on the slope is found from elementary statistics:

$$\sigma_{slope} = \left(\frac{n}{\Delta} \right)^{\frac{1}{2}} \sigma_r \tag{2.13}$$

where σ_r is the noise on an individual sample and

$$\Delta = n \sum_{i=1}^{n} t_i^2 - \left(\sum_{i=1}^{n} t_i \right)^2 \tag{2.14}$$

where t_i is the time of the ith sample. The slope error can then be expressed as

$$\sigma_{slope}^2 = \frac{n \sigma_r^2}{n \delta t^2 \sum i^2 - \delta t^2 \left(\sum i \right)^2} \tag{2.15}$$

where δt is the sampling interval, $(dt = i \times \delta t)$. This reduces to

$$\sigma_{slope} = \frac{\sigma_r}{\delta t \left[n (n+1) \frac{(n-1)}{12} \right]^{\frac{1}{2}}} \tag{2.16}$$

The actual read error, σ_{re} is found from the slope×integration time

$$\sigma_{re} = \sigma_{slope} (n-1) \delta t \tag{2.17}$$

so that, finally the read noise can be expressed quite simply as

$$\sigma_{re} = \frac{\sigma_r}{\sqrt{\left(\frac{n}{12} \right)}} \tag{2.18}$$

$\sqrt{\frac{n}{12}}$

in the limit where n is large. Thus the read–out noise, σ_{re}, is reduced by a factor $\sqrt{\frac{n}{12}}$ compared with that for a single read, σ_r *e.g.* for 1000 samples, read noise would be reduced from $400e^-$ to $44e^-$. The results from using this noise reduction algorithm with CGS4 can be seen by measuring dark currents with and without the NDR mode (Figure 2.11). The number of samples is calculated in the array control electronics as an the closest integral multiple of 63ms to the exposure time requested by the observer; 63ms is the length of time required to read–out the array (43ms) and perform the calculation. Thus a 5s NDR exposure would actually become a 4.977s exposure during which the signal would be sampled 79 times and the resulting read–noise would be $\sim 116e^-$. The reduction in read–noise is more dramatic for the longer exposure times general employed for the astronomical observations, eg for a 40s exposure the read noise would be reduced to $\sim 40e^-$. Figure 2.12 shows a 35s dark observed in NDR mode for which the read–out noise is $34e^-$.

For NIR astronomy, the optimum observations will be made when background limited performance (BLP) is reached, *i.e.* when all other noise sources are dominated by that from the sky and telescope (Section 2.8.1) in the shortest time possible for maximum observing efficiency. For the specific case of the NIR, the variations in the OH emission from the sky place an upper limit on exposure times of ~60s (Chapter 4), so ideally BLP should be reached on shorter timescales than this. It is very simple to show the use of the NDR algorithm greatly improves our chances of achieving BLP. The background noise limit is reached when

$$(Ib\ t)^{\frac{1}{2}} = \frac{\sigma_r}{\sqrt{(\frac{n}{12})}} \tag{2.19}$$

Ib is the background current in photoelectrons (see Section 2.8.1). The number of samples, n, can be expressed as the product of a sampling rate, S, and the exposure time, t, so that the time required to reach BLP, t_b is

$$t_b = \sigma_r \sqrt{\frac{12}{SIb}} \tag{2.20}$$

The sampling rate S is ~15 (1/63ms) with the current CGS4 set–up. Taking a background current of 3000 photons s^{-1}arcsec^{-2} and a read–out noise of $400e^-$, the time to

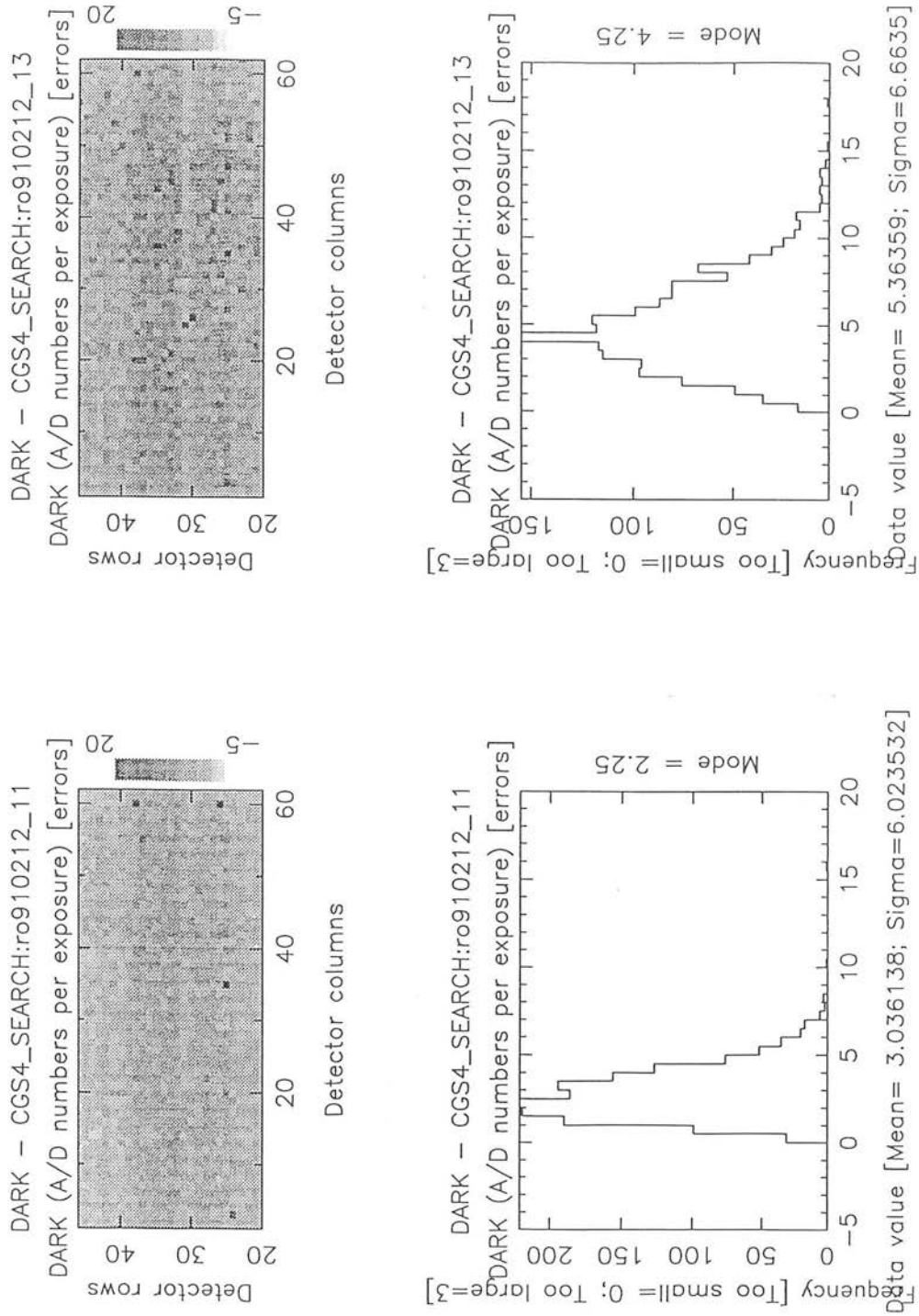


Figure 2.11: A comparison between two darks if the same exposure time, 3s. The theoretical reduction in read–out noise is 2 for the 48 samples; the measured reduction is 1.9.

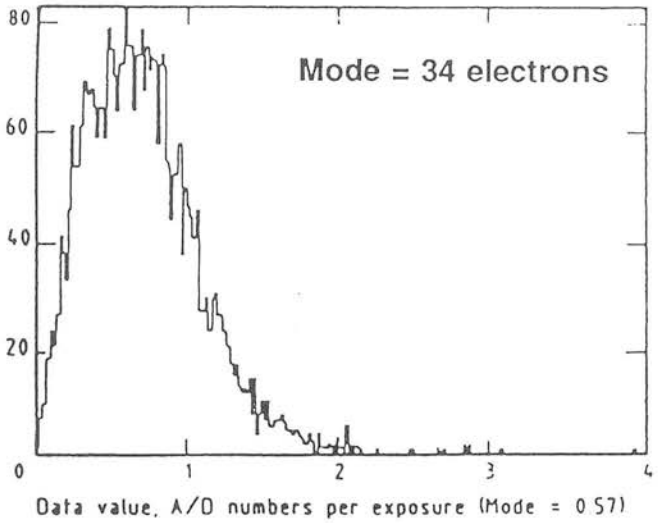


Figure 2.12: The results of using the NDR algorithm with a large number of sample points. *i.e.* for a long exposure time. In this instance, the exposure times was 35s, 583 samples were read and the resultant read–out noise is $34e^-$ (0.57 A/D numbers).

BLP would be 55s with a single read and 7s with NDR.

2.6 Aluminium Optics and the use of Diamond Machining

The 2D array is not the only novel feature in the design of CGS4. Advances in the technique of machining using diamond-headed tools promoted a move away from the traditional methods of optical alignment and design. This instrument has pioneered the use of an all-aluminium construction for the mirrors, support structures and cryostat. The material used is the aluminium alloy 6061T6, which is coated with gold to give 98% reflective mirror surfaces in the NIR. This approach was adopted to satisfy the tight tolerances required to meet the image quality specifications discussed in Section 2.2 when cooling large mirrors ($\sim$ 150mm diameter) to 70K. Using a single material for the entire optical construction eliminates differential contraction on cooling, maintaining the geometrical relationships between the components, the radii of curvature *e t c* . This choice for the construction has many other advantages in attaining the design goals outlined in

Section 2. The use of diamond machining allows the mirror surfaces to be specified as a general asphere. No optical system is perfect, aberrations are always present and can severely affect image quality and instrumental sensitivity, so that most traditional optical systems will comprise several spherical mirrors to minimise aberrations. Another factor is the comparative ease of manufacture of spherical shapes. Using an asphere reduces the number of reflective surfaces in the optical train, thereby reducing the light–losses and aberrations and giving improvements in both image quality and system throughput over a more traditional design. Reducing the number of components inside the cryostat also reduces the thermal mass to be cooled to liquid nitrogen temperatures.

To achieve the flexibility which was required of CGS4, where possible, the elements on the optical train were constructed as independent modules. The modules are: image rotator, filter, cvf/slit and camera/collimator; these are indicated in Figure 2.1. The grating housing can also be considered as a separate module, as can the detector housing/translation mechanism. The two cameras (f/1.35, f/2.70) comprise two sets of mirrors which are housed in different modules, and can be interchanged without need for realignment. The actual method of alignment is detailed in Chapter 3. In essence, the approach adopted is a mechanical one; each mirror is machined with a reference surface or pad. The alignment of the mirror is achieved using laser light (HeNe, λ=632.8μm), with each module aligned separately to the aluminium support structure and a final test of the complete assembly carried out.

2.7 Instrument Control and Automatic Data Reduction

CGS4 is a complex instrument, requiring nine inaccessible motors to be controlled and the data from a 2D array to be obtained and reduced to a scientifically meaningful state. To make the instrument sufficiently straightforward to allow it to be operated as a common–user instrument, sophisticated software systems have been developed to control the configuration of the instrument motors and array controls, the read–out of the array and the reduction of the data. An automatic data reduction system was required to allow real time assessment of the observational data at the telescope. Once this need had been identified, it was evident that a parallel system to provide a highly structured

mode of observing would also be required, in order that the data reduction system would correctly handle the observations and their corresponding calibration frames. The data and acquisition system was developed in such a way as to allow the observer to carry out detailed sets of observations, such as mapping of extended sources, in a straightforward way.

The automatic data reduction software, CGS4DR, is based on FIGARO subroutines, and operates on a menu–driven system. It has the capability to produce and display fully reduced, bias and dark subtracted, flat–fielded, sky–subtracted spectra in real–time. Other applications include the ability to perform simple arithmetic functions on the frames and to fit lines in the spectra automatically; these are not widely use during normal operation of the spectrometer, but were invaluable for the engineering tests. A wide range of plotting options is available; many of the plots in this Chapter were produced using CGS4DR. The importance of the automatic data–reduction facility for instant assessment of both the calibration and astronomical observations, allowing quick reactions to their quality and greatly improving the use of observing time, cannot be overstated. CGS4DR was widely used during the laboratory commissioning of the instrument, which speeded up the analysis of the experimental data as well as thoroughly exercising the software.

All of the software systems operate under menu–driven environments. Additionally the array control and acquisition systems can be operated using macros called EXECs, which are text files of recognised commands created by the observer. These operate in an automatic way, and will perform a series of observations, resetting the motor configuration, the array exposure time *etc* and taking observations of calibration frames as well as object and sky exposures. A data reduction configuration can be loaded into CGS4DR using files called by an EXEC. The data acquisition software can also communicate with the telescope, so that telescope offsets may be contained in the EXECs. The EXEC can be created prior to going to the telescope, greatly improving the efficiency with which observing time is deployed. Since the telescope offsets can be defined in an EXEC, they provide a useful facility for carrying out mapping experiments. The EXEC system was used for the engineering tests on the instrument, reported in Chapter 3, for example, repeatability and reproducibility tests involving many changes of instrument configura-

```
! february 10th
! exec for h2 emission from the orion bar
! in the k window
config orion_bar_k      ! load configuration ``orion_bar_k'' into software
drconfig orion_red      ! adopt configuration ``orion_red'' for CGS4DR
bias                    ! set motor positions and make Bias observation
dark                    ! set motor positions and take Dark observation
set object              ! set motor positions for object
name the orion bar      ! set object name
targetwait orion bar    ! move telescope to defined position
break                   ! pause for ``peaking up'' on source
startgroup              ! flag the start of a new set of observations
do 5 quads              ! repeat the exec QUADS five times
endgroup
```

```
! exec quads
object      ! take object observation
tel offset  ! move telescope to defined sky position
sky         ! take sky observation
sky
tel main
object      ! return telescope to defined object position
```

Figure 2.13: Two Simple EXECs which will perform a complete set of observations of an extended source. The EXEC orion_bar.exec calls the second EXEC, quads.exec.

tion could be performed overnight. Figure 2.13 shows two simple execs which carry out a complete set of observations of an extended source. The instrument configuration (the motor positions, integration times etc.) is set by the CONFIG file, which is read by the EXEC. An example CONFIG file is shown in Figure 2.14.

Another useful feature of the acquisition software is the facility for taking quick snapshots of the signal on the array; the exposure times for this can be a few tenths of seconds. This facility is referred to as MOVIE. At the telescope its primary use is for centering guide stars on the slit by maximising the observed signal in a row. It is also used for checking for saturation of the array, *e.g.* when taking flats, or for monitoring the array behaviour during use.

```
Astronomical     configuration
Basic (object) configuration:
ND_STARE              CIRACS configuration
15.0                  exposure time
     1                exposures/integ
     1                scans
     3                sampling
1_pixel                  sample range
45                    position angle
B2                    filter
1_pixel               slit width
75_lpmm                 grating
2.2                   wavelength
 1                    order
off                   calibration lamp
  100                 tungsten-halogen level
  100                 lamp effective aperture
Flat configuration variant (from object):
     1                exposures/integ
     1                integrations
Dark configuration variant (from object):
     3                exposures/integ
     3                integrations
Bias configuration variant (from dark):
0.043                 exposure time
  1000                exposures/integ
     3                integrations
```

Figure 2.14: The CONFIG file containing the motor positions and exposure times which will be used in the observations controlled by the EXEC star.

2.8 Predicted Instrumental Sensitivities

By assessing the likely noise contributions to an observation and by making certain initial assumptions about the system throughput, the sensitivity of a new instrument can be estimated.

2.8.1 Sources of Noise in the Near Infrared

NIR astronomy is plagued by noise sources, each of which will dominate in a particular wavelength range, for a given object strength or exposure time. The photon noise from the object itself is governed by Poisson statistics and calculated as $\sqrt{\text{detected photons}}$ from the source. This is the dominant noise source for only the brightest objects, such as calibration stars, rarely for the object of scientific interest. Possibly the thing which truly distinguishes infrared astronomy from astronomy in the visible waveband is the strong background signal from sky and telescope (Figure 2.15).

This predicates the way in which we observe, either nodding or chopping to remove the sky–signal, and is often an order of magnitude stronger than the object signal and so dominates the noise. The sky background is sub–divided into two distinct regimes:

(*i*) **1–2.3μm** In this regime the sky background emission is dominated by the non–thermal emission from OH radicals in the upper atmosphere, giving a spectrally structured and variable source of noise. Beyond about 2.0μm the thermal background component will start to increase. The strongest OH lines produce ~1,000s of photons per second per telescope area, with intensity variations of typically 10% over periods of ten minutes, restricting exposure times for the object observations to ~60s to prevent residual sky signal in the sky–subtracted object spectrum contributing to the noise. Details of the variations of OH emission, the mechanism which causes them, and the noise contribution to an observation are given in Chapter 4. This noise source is most troublesome when observing with the low–resolution mode with many OH lines present in the spectral region observed. When observing with the echelle, the separation of the OH lines is such that there may

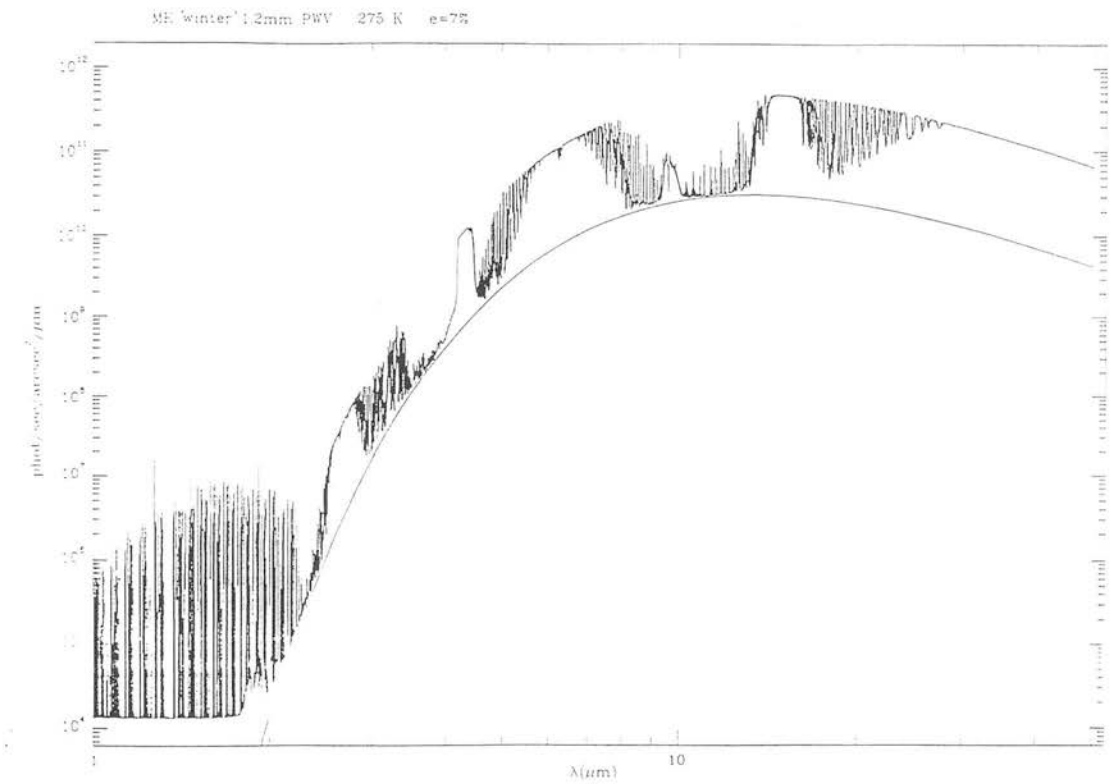


Figure 2.15: The emission from the sky at 1–5μm . The emission below 2.3μm is line emission from OH in the upper atmosphere; above 2.3μm continuum emission from the sky and telescope dominates.

be no lines within the wavelength region observed, giving a very low–background regime and high sensitivity.

(*ii*) **2.3–5.0μm** At longer wavelengths the sky background is dominated by thermal emission from the telescope and sky which radiate as a ~270K grey–body.

The third noise source is detector noise, which has been described in Section 2.5. In general, there will also be some contribution to the dark current from "light leaks" and thermal emission from the dewar. This effect is negligible with CGS4.

2.8.2 Estimating the Sensitivity for CGS4

To estimate the instrumental sensitivity, the total noise contribution for a given wavelength of observation with a given telescope and instrument/detector must first be assessed. The parameters describing the telescope/instrument/observation configuration are as follows.

A number of fixed parameters describe the optical components; these are assumed to be known and constant:

telescope emissivity, ϵ_t=0.1
telescope throughput: η_t=1-ϵ_t
telescope diameter, D_t=3.8m
CGS4 grating efficiencies for the low–dispersion gratings, η_g are listed in Table 2.2. These are supplied by the manufacturer, Milton Roy.
pixel size, D = 76μm
the angle off–axis which the beam makes with the grating, $\alpha = 14°$.
the array size in pixels along the dispersion direction, m=62

There are free parameters associated with the instrument which are "best–guesses" before the instrument is built and tested.

dark current, $I_d = 1.5e^-s^{-1}$ from 1–5μm
read–out noise, N_e=40e^-

Grating Efficiencies			
λ	order	Efficiency, η_g	
(μm)		75 l/mm grating	150 l/mm grating
1.25	1	0.00	0.00
	2	0.75	0.00
	3	0.00	0.83
1.65	1	0.45	0.00
	2	0.38	0.75
	3	0.00	0.10
2.20	1	0.84	0.00
	2	0.00	0.57
3.80	1	0.55	0.95
4.90	1	0.35	0.75

Table 2.2: Grating Efficiencies for the two low–dispersion gratings, as used in the derivation of the instrumental sensitivity.

CGS4 throughput, including quantum efficiency (0.7), but excluding grating efficiencies: η_s=0.17

A further set of parameters describes the instrumental configuration used for a given observation, for example

grating lines per mm, g=75
grating order, n=1
camera focal length, f=150mm
the total observation time, T=30mins
the number of integrations in time T, Ints=10
the oversampling factor S=3
on–chip exposure time=60s
the signal/noise ratio required, σ=3

The noise from the sky background can be found taking the contributions from thermal and non–thermal emission in three parts, using the following approximation.

λ < 2.0um, the emission is entirely non–thermal emission from OH lines.
2.0um< λ <2.4μm , the emission is mixed, 50% non–thermal and 50% thermal. Note that the thermal emission is dependent in the telescope emissivity ϵ.
λ > 2.4μm , the emission is totally thermal.

The background emission for a given waveband is then:

$$sky\,(\lambda) = \eta_t background\,(\lambda) \qquad \text{for} \quad \lambda < 2.0\mu\text{m} \tag{2.21}$$

$$sky\,(\lambda) = background\,(\lambda)\,\epsilon_t \qquad \text{for} \quad \lambda > 2.5\mu\text{m} \tag{2.22}$$

$$sky\,(\lambda) = background\,(\lambda)\,\frac{\eta_t}{2} + background\,(\lambda)\,\frac{\epsilon_t}{2} \qquad \text{elsewhere} \tag{2.23}$$

measured in units of photons s^{-1} m^{-2} $arcsec^{-2}$ μm $^{-1}$. This is then converted into a background current in electrons per second, Ib(λ):

$$Ib\,(\lambda) = sky\,(\lambda)\,A_t FOV^2 \eta_s \eta_g \delta\lambda \tag{2.24}$$

where A_t is the telescope area, FOV is the field of view given by

$$FOV = 23.29 \frac{D}{D_t f} \tag{2.25}$$

All other terms are defined above, except $\delta\lambda$, the dispersion per pixel which is expressed

$$\delta\lambda = \frac{\lambda}{R} \tag{2.26}$$

The resolving power, R, must be found for the particular grating/camera configuration. The calculation proceeds slightly differently for the echelle and the low–dispersion gratings. For the 75 l/mm and 150l/mm gratings :

$$R = 2\frac{f}{D} \tan\left(\theta_g\right) \frac{1}{p} \tag{2.27}$$

and θ_g is the grating angle obtained from the standard grating equation

$$n\lambda = 2\frac{1}{g} \sin\left(\theta_g\right) \cos\left(\gamma\right) \tag{2.28}$$

where γ is the out–of–plane angle,= 14°. For the echelle observations, maximum grating efficiency is attained with all observations made with the grating angle as close as possible to the blaze angle and the spectral resolution is then given by

$$R = 2\frac{f}{D} \tan\left(\theta_{blaze}\right) \frac{1}{p}. \tag{2.29}$$

The blaze angle for CGS4 is 63.5° and the resolution R~8,000 (δV=38km s^{-1} per pixel).

Now the contributions to the noise expressed in photoelectrons may be added in quadrature to give the total noise input for the specified observation

$$N = \sqrt{N_e^2 + (\frac{T}{SObs}[Id + Ib])} \tag{2.30}$$

To relate the calculation to an astrophysically significant scale, the noise can be compared with the flux from Vega in a passband $\delta\lambda$ for a telescope of area A_t and throughput η_t (Mountain *et al.* 1985). The stellar flux is assumed to cover p^2 pixels, and the spectrum result from the addition of p rows. This flux is referred to as the zero point, $Zp(\lambda)$.

$$Zp(\lambda) = 9.8863 \times 10^{10} A_t \eta_t \eta_s \eta_g \delta\lambda \frac{1}{\lambda^4 \left(\exp \frac{1.4388}{\lambda} - 1\right)} \frac{1}{p^2} \tag{2.31}$$

so that the line flux detectable at the σ level specified is given by

$$L = \frac{\sigma}{\sqrt{p}} N \frac{S}{T} \sqrt{Obs} \frac{hc}{\lambda} \frac{1}{A_t} \frac{1}{\eta_t \eta_s \eta_g} \tag{2.32}$$

in W m^{-2} μm $^{-1}$and the magnitude limit will be given by

$$M = -2.5 \log \frac{\sigma}{\sqrt{p}} N \frac{S}{T Zp(\lambda)} \sqrt{Obs} \tag{2.33}$$

The graph in Figure 2.16 shows the 3σ30min magnitude limits for an observation where the stellar image is over 1 pixel only and the observation consists of n samples.

2.9 Measured Sensitivities

The actual sensitivity achieved with the instrument can be calculated by examining the noise on background limited observations of standard stars. The measured magnitude limits for several standards, observed with different instrumental configurations, are given in Table 2.3; the 3σ30min limits have been extrapolated from the length of exposure time of the observation. The measured limits are less then the limits predicted. The best agreement is obtained for the L band observations, where the difference in magnitude is

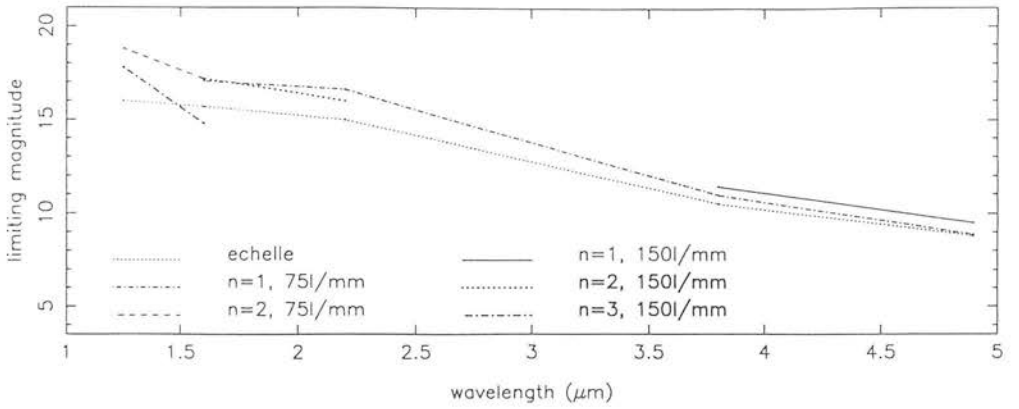


Figure 2.16: Predicted instrumental sensitivities for the low dispersion gratings and the echelle for a 3σ detection of a point–like source, where the image is spread over one pixel. The observation time is 30min.

0.3. For the M band observations, the difference in magnitude of 0.7 implies that the error bars are a factor of two larger an expected. This consistent with observations in which half of the stellar flux is lost due to motion of the star relative to the $3\widehat{''}$ slit. This is one of greatest sources of loss of sensitivity for observations of compact sources with CGS4. The K–band sensitivity is 1.4 magnitudes brighter for 3σ30min observations than expected, *i.e.* the error bar is five times too large. The OH emission variation described in Chapter 4 could be the cause of such high noise levels, as this is the wavelength region in which it dominates. The on–chip exposure time for this star was 20s.

2.10 Calibration Observations

2.10.1 Detector Effects: Bias and Dark Observations

An integral part of obtaining the correct flux calibration is removing the detector darks and biases from the observations. The bias voltage of -250mV is applied to the array, giving a well depth of $\sim 8\times10^5$ electrons. The measured bias level is therefore $\sim 3.3\times10^4$ counts (A/D numbers) in the array output analog/digital convertor (ADC), since the

Measured Sensitivities for Observations with CGS4			
standard star	instrumental configuration (grating, order, window)	3σ30min magnitude limits	
		measured	predicted
HD129655	75 l/mm, 1, K	16.1	17.5
HD161903	75 l/mm, 1, L	11.5	11.8
BS7615	75 l/mm, 1, M	9.1	9.8

Table 2.3: The measured magnitude limits of 3σ30min observations are compared with those modelled using the method outlined in Section 2.8.2.

gain for the CGS4 array is $24e^-$ per A/D number. To maintain the linearity of the array response, the wells should only be filled to 80% well during an observation. The ADC will saturate at a level of 2^{16} (65536) A/D numbers; a voltage offset applied to the array will ensure that this saturation point is reached before the wells are 80% filled. DC voltage offsets in the read–out electronics causes each pixel to assume different voltage level, shown by the bias structure (Figure 2.17). These offsets must be determined and removed from the observations to calibrate the array.

A bias frame is a short observation of the dark cryostat; the optical path to the window is blanked off so that no light is incident on the array. The exposure time is set to near the minimum time for reading out the array (43ms) so that only the voltages from the array and electronics are measured; if a longer exposure time was used the bias structure would be measured but with the addition of the dark current accumulated during the exposure. It is assumed that the dark current is negligible in 43ms, a fact which is verified by the measured dark currents. The stability of the bias observations is an important consideration, as they may be used to calibrate observations taken several hours apart. Figure 2.18 shows the modal bias level for several biases taken during the night of 28 February 1991. It can be seen that the variation in the level during that night is $\sim$4%. The stability of the dark frames will also be represented by this figure, as the dark current is negligible ($1e^-s^{-1}$) compared with the bias level.

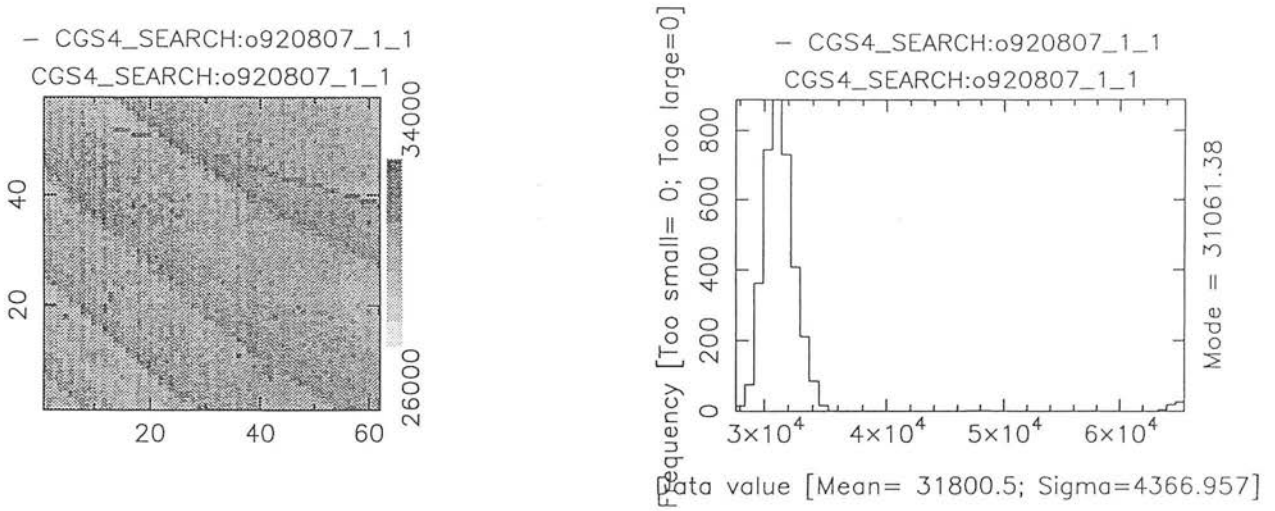


Figure 2.17: A typical bias from for the CGS4 array.

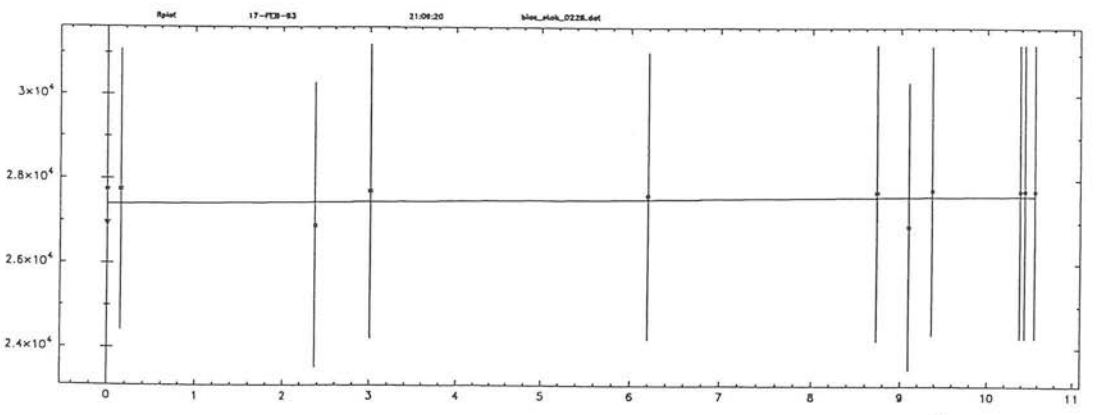


Figure 2.18: The modal level for a series of bias frames taken during one night of observations. The variation in the level is ~4%.

A more significant dark current will be detected during a longer observation of an astronomical source, particularly with faint sources when integrations times will be tens of seconds. Since dark current increases with time, a dark with the identical exposure time to the object frame is required to subtract the dark current from the object signal. Moreover, the dark and object should both be observed in the same mode, either STARE or NDR, due to heating effects with multiple sampling which would increase the dark current. The instrumental configuration for the dark is the same as that for the bias.

In fact, the steps outlined above are redundant for many observations types in NIR astronomy, since the complete reduced observation consists of an object frame with a sky frame subtracted from it. The bias levels and dark current will be the same for both observations of the pair and therefore the bias and dark are effectively subtracted. This has a signal/noise advantage: if the dark frame and bias frame are subtracted from both the object and sky frames then the variances on each are added in quadrature giving excess noise over the object–sky frame

$$\sqrt{(2\sigma_d^2 + 2\sigma_b^2)}\,. \tag{2.34}$$

The automatic data–reduction, CGS4DR, will remove the bias and dark observations if the observer so chooses. The particular dark frame to be used may be selected by the software by searching for the nearest (in time) suitable dark; the only check of suitability is that it should have the same exposure time as the object. Alternatively, the observer can specify a dark that should be used.

2.10.2 Bad pixel masks

A raw image of the array shows many cosmetic imperfections. These are the "bad pixels" which may have a lower QE than the others, have no response to incoming photons at all or be permanently saturated or "warm", *i.e.* appearing to have greater QE than the rest of the pixels. The signal from these pixels, except those with no signal or a fully saturated signal, is variable in time and the bad pixels are manifest by showing greater errors than their neighbours. To ignore these pixels, a mask can be created which flags

the values from them as "bad", and thereafter they are not used by the data reduction system. The map of the positions of the bad pixels is referred to as a bad pixel mask (BPM); it is applied to each integration creating the so–called quality array for that observation.

The bad pixels can be identified by the size of their errors relative to those on the normal pixels. A histogram of the error distribution from a dark frame, for example, will show a gaussian distribution of errors with the bad pixels showing values many standard deviations from the mean. The data reduction system can be used to identify these pixels by thresholding the data frame, thereby creating the BPM. For most configurations used with CGS4, it is also necessary to flag the pixels in the unilluminated area of the array as bad. A separate "window" mask can be defined by the extent of the illuminated area, and then the threshold and window masks combined. This is particularly critical when normalising the flat–field frame, as the unilluminated pixel values would otherwise distort the mean value (see Section 2.12.1).

By creating the bad–pixel mask from a frame which has been sampled by stepping the detector array over two pixels (see Section 3.11), most of the bad pixels are effectively removed. Figure 2.19 shows two bad pixel masks created for the array FPA180, the first from a observation which was sampled over a single pixel, the second from an observation which was sampled over two pixels. The pixels which remain are those at the edge of the array or where two or more bad pixels are adjacent.

2.11 Flat Fielding

A flat–field defines the relative responses of each individual pixel of the detector array, which may be referred to as the gain matrix of the array. The gain variations arise since each pixel is read through a separate electronic path and will also have a different quantum efficiency from its neighbours (Section 2.5). Flat–fielding is the process of correcting for this through division by a flat–field. There are many reasons why this process is an essential part of the calibration of astronomical data. Pixel–pixel values must be corrected in the dispersion direction for accurate determination of spectral line

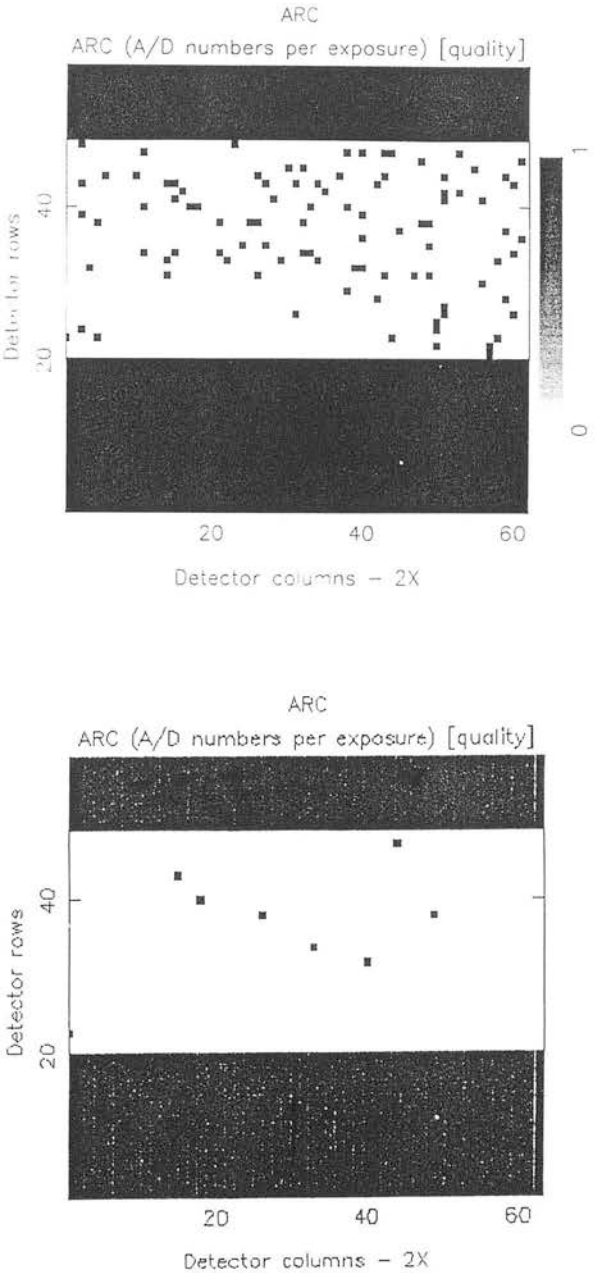


Figure 2.19: Two bad pixels masks for FPA180. The leftmost was created from a frame which was not oversampled, whereas that on the right was created from a frame which was sampled over two pixels, effectively removing all bad pixels except where two occurred adjacently or the bad pixel occurred at one side of the array. The frames have been combined with a window mask appropriate to the illuminated area seen with the low–dispersion gratings.

ratios from either extended or compact sources. For extended sources, the changing pixel gains along the slit must be removed for comparison of emission from different spatial points in the source or so that a number of rows may be coadded without introducing a systematic error. For point–like sources, removal of the residual sky lines by fitting along the slit and extrapolating to determine the residual component of the object signal assumes that the spatial field is flat over the $90\widehat{\prime\prime}$ slit.

The requirements for a good flat–field source are that the light from the source pass along the same optical path as the light from the astronomical source and therefore be subject to the same vignetting *e.t.c.*, the illumination should be uniform both spatially and spectrally over the entire array and the emission should be stable both over long timescales, so that flat–fields can be taken at dusk and used to reduce the night's observations, and over short timescales, so that the errors in determining the gains are minimised. The accuracy of the gain measurement is also critical, as it affects the accuracy with which the sky frame can be removed and the final signal/noise of the observation. Typically, for the signal/noise on the flat–field observation to contribute only 1% of the error on a 10σ observation of a source, where the object strength is 1% of the sky signal (10,000 photons s^{-1}), the flat field will have to be accurate to $1{:}10^4$.

The pixel responses can be measured by illuminating the array with a uniform source of light fed through the optical system. This image is termed the flat–field by which observations are divided to remove the pixel–to–pixel variations. In practice, for near–infrared spectroscopy achieving an adequate flat–field to correct for these variations both spatially and spectrally is far from trivial. To a great extent, this problem is peculiar to 2D array detectors in the NIR; for previous generations of linear array detectors (*e.g.* CGS2) the relative pixel gains were removed by ratioing the object with a standard star. To flat–field a 2D array in this way would require a star on every row, clearly not a practical proposition.

Techniques for producing a flat–field image are well developed for optical astronomy and infrared imaging. For the latter, it is conventional to observe blank areas of sky, with a small spatial offset introduced between each. The frames are then coadded and the median of all the frames found, removing any stars in the frames. However, it is not

possible to use the sky to flat–field a spectrum because of the wealth of spectral structure provided by the emission from OH molecules in the upper atmosphere (Chapter 4) and by the atmospheric transmission profiles. Another possible solution is to illuminate a screen or area of the telescope dome with a bright fluorescent light, producing a uniformly illuminated area. This technique is unsuitable for NIR spectroscopy for two reasons: for the 1–2.5μm region, the dome surface will not act as a diffuse scatterer (see Section 2.11.1) and therefore the field will not be uniform. For the thermal infrared, it is the black–body emission from the 275K dome which will provide the flat, and it is inadequate both in terms of the observable photon flux and the temperature stability over long and short timescales. For a 275K black–body the change in temperature which will cause a change in the black–body flux of 1:10^4, the required accuracy for the flat–field, is 1.57mK.

To provide a uniform field for flat–fielding CGS4, a dedicated facility, the "calibration unit" (CU), was designed. It houses two continuum sources for flat–fielding, a black–body source and a tungsten–halogen (t–h) lamp (Section 2.11.1). and a set of gas discharge lamps for wavelength calibration. The optics of the calibration unit re–image the output from the calibration sources to provide a field identical to that of the telescope *i.e.* an f/36 beam which is brought to a focus at the slit and which fills the Lyot stop in the same way as an image of the telescope secondary does.

2.11.1 The calibration unit

Given the stated difficulties of obtaining a flat–fielding source for NIR spectroscopy (Section 2.11), such a source can be provided in a carefully specified calibration unit by including continuum sources and ensuring uniform illumination of a simulated f/36 beam like that of UKIRT. By filling the 90$\widehat{''}$ field with the same beam as the telescope, we approach the first condition for producing a good flat–fielding source. Although the telescope optics are not included in the optical train of the flat–field, the optics of the calibration unit reproduce the illumination of the Lyot stop as it would be for the telescope secondary. The signal/noise considerations are fulfilled by the continuum sources, as outlined in Section 2.11.3. The question of obtaining a uniform illumination

over the entire array is addressed by the use of an integrating sphere with a perfectly diffusing surface. The theory and practice of such a device in described in Section 2.11.2.

The operation of the CU is controlled via the software which selects the instrument configuration. When the calibration unit is not in use, the light from the telescope passes through it unobstructed into the cryostat; when it is selected a mirror moves into the telescope beam which reflects the output from the calibration lamps and flat–field sources into the cryostat. Three arc lamps, argon, xenon and krypton, are available in the CU; standard wavelengths adopted for these lines are from Outred (1978). The three lamps provide sufficient lines to give a low–order polynomial fit to spectra for the low–dispersion gratings, although providing adequate lines to calibrate the echelle is more of a problem. Typically an arc spectrum with 6 lines will give a fit accurate to 0.1–0.2 pixel. For the echelle, as the wavelength coverage is only ~0.002um across the entire array, a configuration may be chosen in which no arc lines can be observed. In this case, if $\lambda < 2.3\mu$m the sky emission lines can be used as calibration sources, but the accuracy to which their wavelengths have been determined is ~15km s^{-1} or half a pixel. (Oliva & Origlia (1992) give a list of observed OH lines with wavelengths accurate to 0.5Å). Thus, for the echelle, OH lines should not be considered a standard for wavelength calibration, but as a useful backup. The continuum sources chosen are a 1173K black–body and a tungsten–halogen (t–h) lamp. The choice of these sources is described in Section 2.11.3.

2.11.2 Integrating Spheres and Lambertian Surfaces

Integrating spheres are routinely used for applications where a uniform source of illumination is required. A uniform source is one for which the illumination is not dependent upon the angle from which it is viewed; such a source may be referred to as a Lambertian source. An approximate description of their operation is as follows: light enters from an input port (of which there may be several around the sphere) and is reflected from the inner surface of the sphere; some fraction of the light is emitted at the designated exit port. The uniformity of the output field is achieved by the multiple reflections from the roughened inner surface of the sphere which acts as a diffuse scatterer, effectively erasing any memory of the input source so that, for example, no image of the filament of a bulb

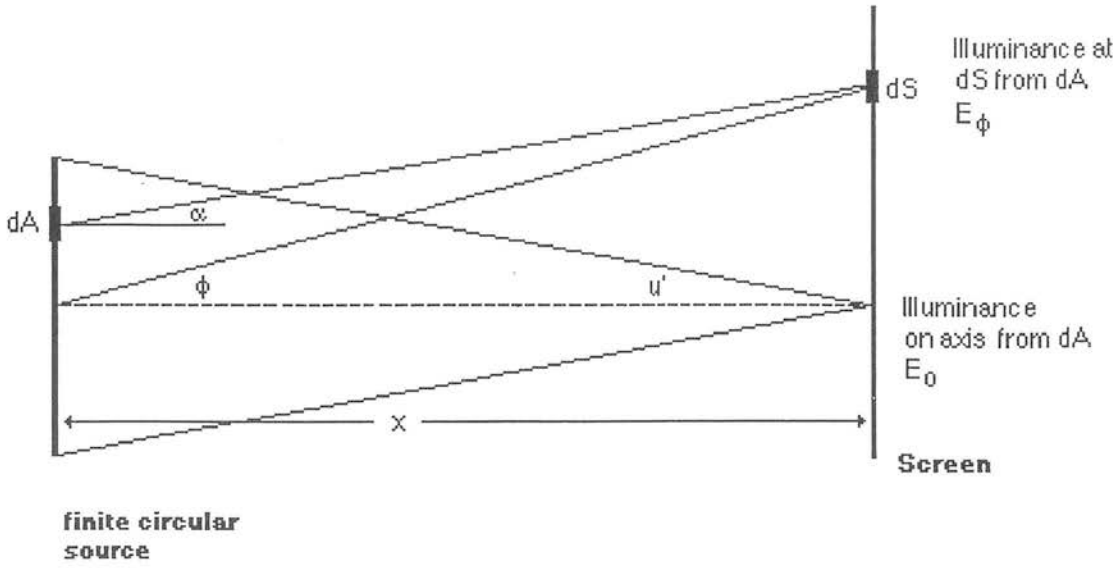


Figure 2.20: The illumination from a finite, circular source at a screen, at a distance, x. The relationship between the on–axis illuminance, E_o, and the illuminance on an area, dS, at an angle, ϕ, determines the uniformity of the illumination.

would be seen at the exit port. Often, baffles are included inside the sphere to prevent the exit port having a direct view of the input source. This uniformly illuminated, circular output port can be treated as a finite, circular source, and the strength and uniformity of the irradiance of a target by such a source can be calculated (Kingslake 1965). A schematic diagram of the optical set–up under consideration is shown in Figure 2.20. For an area of the target illuminated by an on–axis beam the irradiance is given by

$$E_o = \pi B \sin^2 U' \tag{2.35}$$

where B is the radiance of the source, and U' is the half–angle subtended by the source at the target. For an off-axis angle, ϕ the irradiance becomes

$$E_\phi = \pi \sin^2 U' \cos^4 \phi [F'] \tag{2.36}$$

where ϕ is the half–angle subtended by the detector at the source, and $[F']$ is referred to as the increase factor or the configuration factor and is specific to the particular choice

of target and output port diameter. For small values of U' and ϕ', $[F']$ ~1 and E_ϕ can be approximated by

$$E_\phi = E_o \cos^4 \phi \tag{2.37}$$

The ratio of E_ϕ to E_o is the spatial uniformity of the irradiance. This sets an upper limit to the output port diameter where high uniformity is required. The opposing factor, which seeks to increase this diameter, is the need for high transmission. The transmission is given by:

$$\tau = \frac{\phi_e}{\phi_i} = \frac{\rho f_e}{1 - \rho(1 - f_j)} \tag{2.38}$$

assuming that at the first reflection all of the light from the source strikes the diffusing surface. The fluxes from the exit and input ports are ϕ_e and ϕ_i respectively, ρ is the reflectance of the surface, f_e is the area of the exit port, as a fraction of the total surface area of the sphere, and f_j is the sum of all the port areas as a fraction of the total surface area, which should be restricted to ~5% to ensure uniformity over the field.

The integrating sphere used in CGS4 was manufactured in the ROE workshops; the design plans are shown in Figure 2.21. There are two input ports behind which the t–h lamp and the black–body are located; the output port has no view of the input ports so no baffling was required. The input ports are both 6mm in diameter, the output port has a diameter of 5mm and the diameter of the sphere is 30mm. Thus the percentage of the total surface area occupied, for f_e=0.0069 and f_j=0.0269, is only 2.7%, giving a theoretical uniformity of 99.5%. The CGS4 sphere was treated by Labsphere with InfraGold, a proprietary coating which acts as a Lambertian scatterer for the NIR. Clearly the reflectance is also an important factor in optimising the transmission: the roughened surface of the sphere is gold–coated giving 95% reflectance at 1–5μm . The theoretical transmission, τ, is 0.146. A previous attempt to produce an integrating sphere using a titanium–oxide coated surface failed, highlighting the problems associated in finding a diffuse scatterer for the NIR.

The CGS4 integrating sphere is expected to produce a field with a uniformity over the $90\widehat{\text{m}}$ of 99.5%. If this field is subject to spatial vignetting due to the CGS4 optics,

then the illumination of the detector will not be flat along the slit, as required. This can only be calibrated using a secondary flat source. A method of doing this using the sky is outlined in Section 2.12.4.

2.11.3 Choice of Continuum Sources

A clear choice of continuum source for the CU, to cover the signal/noise requirements, would be a black–body of peak wavelength close the centre of the 1–5μm range of the instrument. Choosing 2.5μm as the peak wavelength implies a black–body of 900°C. The black–body source should have sufficient flux to give a high signal/noise ratio in modest observation times. The arrival rate of photons from the black–body at the detector can be simply calculated. The radiance of a black–body is given by

$$B_\lambda = \frac{2hc^2\lambda^{-5}}{\exp(\frac{hc}{\lambda k T_{bb}}) - 1} \tag{2.39}$$

in units of $\mathrm{Wm^{-2}sr^{-1}}$ per unit wavelength interval, where h is Planck's constant, c is the speed of light , k is Boltzmann's constant, λ is the wavelength of the photons and T_{bb} is the temperature of the black–body source. For a 900°C black–body, the flux of photons per unit wavelength at 2.5μm is 9.0×10^9 $\mathrm{Wm^{-2}sr^{-1}}$. To preserve the uniformity of the field, the semi–angle subtended by the input port of the integrating sphere is 5°, so that the flux transmitted through the sphere and emitted at the output port is

$$B_\phi = B_\lambda \phi^2 \tau. \tag{2.40}$$

The transmission τ of the CGS4 integrating sphere is 0.146 (from Equation 2.38). Thus the flux emitted from the output port and the photon arrival rate at the detector are given in Table 2.3.

Since the AΩ product is conserved throughout the system, and the exit port of the sphere and the detector plane are conjugate planes, the flux from the output port is the same as that incident upon the detector before losses are taken into account. The solid angle subtended by the detector is defined by the f/# of the camera and the

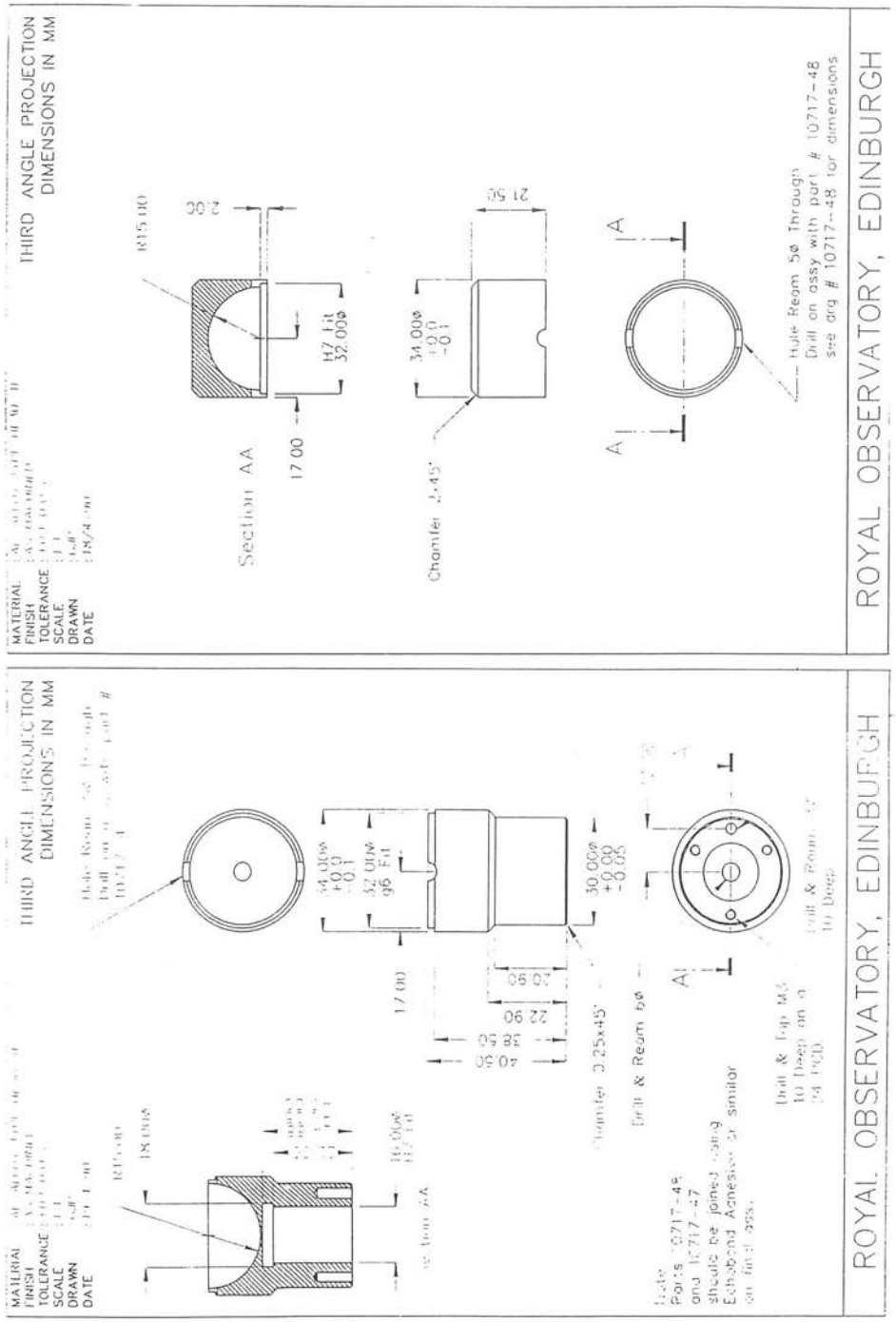


Figure 2.21: Technical drawing of the CGS4 integrating sphere.

Photon Arrival Rates for the Integrating Sphere.		
λ (μm)	75 l/mm grating	echelle
1.25	1.25×10^{6}	9.68×10^{5}
1.65	5.37×10^{7}	3.18×10^{6}
2.20	2.24×10^{8}	7.23×10^{6}
3.80	1.66×10^{8}	8.48×10^{6}
4.90	7.64×10^{7}	6.28×10^{6}

Table 2.4: Photon arrival rate (photons per second per pixel) for observations of a 1193K Black–body with the CGS4 integrating sphere. These figures assume the short focal length camera; an approximate value for the long focal length camera arrival rates will be obtained by dividing by 8, due to the difference in field of view of the lflc ($\times$ $\sim$4) and the change in resolution ($\times$2).

detector area. The photon rate has been calculated for realistic configurations of the grating; the grating efficiency has been included (Table 2.2), and a total transmission for the spectrometer, including integrating sphere losses, is taken as 0.15. The values tabulated are for the 150mm focal length camera; using the 300mm focal length camera reduces these numbers by a factor of $\sim$8 because of the change in both resolution and field of view. From the photon arrival rates in Table 2.3, it is seen that the well depth of $\sim8^5$ electrons may become filled within the minimum read–out time for the array (43ms). A variable aperture was fitted for the black–body, under motor control, to prevent this occurance. For all but the shortest wavelengths, the signal is sufficient to give a signal/noise of $\sim$1,000 for a single exposure. However, to ensure this accuracy at the shortest wavelengths also a tungsten–halogen lamp (colour temperature 3000K) was also purchased. The t–h lamp may also be used at wavelengths where the black–body would saturate, although for the long wavelengths there are few photons from the lamp.

2.12 Flat–fielding Methods with CGS4

In the following sections each frame is considered to be an array(i,j) of pixels arranged in N_i columns parallel to the slit and N_j rows running in the dispersion direction. Therefore the index i runs along the wavelength direction.

2.12.1 Pixel–Pixel Gain Variations in the Dispersion Direction

The requirements for producing a flat–field for NIR astronomy were summarised in Section 2.11. The following is the scheme adopted for determining the pixel–pixel gain variations in the dispersion direction for CGS4 observations.

CGS4's calibration unit has been described in Section 2.11.1 above. Its importance in fulfilling the conditions for an accurate flat–field is paramount. The two continuum sources provide uniform illumination are a black–body at 1193K located at one port of the integrating sphere or, for shorter wavelengths, a tungsten–halogen bulb. The output of the calibration unit is a highly uniform f/36 beam, matched to the beam size of UKIRT and focused at the slit position. In this way, the best approximation to an identical optical path for the flat–field and the astronomical sources is found, given that we are unable to use the sky itself. The observed frame will have individual pixel levels which can be described:

$$lamp(i,j) = B_\lambda(i) * I_{bb}(j) * W(i) * V(j) * G(i,j) \tag{2.41}$$

where B_λ is the black-body spectrum from the CU, I_{bb}(j) describes the intrinsic variation in intensity of this spectrum along the slit (a slowly varying function determined by the output of the CU) W(i) is the normalised instrumental profile as a function of wavelength (the result of the grating optics, broad-band filters etc. but not the atmosphere), V(j) is the normalised instrumental profile as a function of position along the slit (eg. as a result of vignetting in the system) and G(i,j) is the normalised array relative pixel-pixel gains as a function of position – the "gain matrix".

By averaging over the rows, i.e. over j, we obtain a 1D spectrum $lamp\_row\_avg(i)$ which represents the instrumental response as a function of wavelength:

$$lamp\_row\_avg(i) = B_\lambda(i) * W(i) * < I_{bb}(j) * V(j) * G(i,j) >_j \tag{2.42}$$

where the terms in $< >_j$ are averaged over rows j.

The resulting form of this spectrum shows low–frequency variation which represents B_λ*W(i) which is fitted by a low–order (3–7) polynomial $lamp\_row\_pol(i)$, and high frequency variations which are the $<\ldots G(i,j)>_j$ terms, *i.e.* the pixel–pixel variations. Dividing lamp(i,j) by lamp_row_pol(i,j) row-by-row will then give us a "partial flat field" FF1(i,j) where the relative gains within each row are correct, but those along each column j are not, due to the terms V(j) and I_{bb}(j).

$$FF1(i,j) = I_{bb}(j) * V(j) * G(i,j) * constant. \tag{2.43}$$

The above is the algorithm used by the automatic data reduction system for producing the flat–field. An observation of the continuum source is made at the same wavelength and with the same configuration as the object. The order of polynomial to be used to fit the spectrum formed by collapsing the image along the slit is specified by the observer. The resulting polynomial, $lamp\_row\_pol(i)$, is divided into every row of the flat–field frame giving the pixel–pixel response normalised to unity. There is also the option not to normalise the flat–field frame. The resulting frame is then applied to each individual integration, as the flat-field is not over–sampled. The result of flat-fielding an arc lamp spectrum is shown in Figure 2.22, which shows two arc lamp spectra, one flat–fielded, the other not.

2.12.2 Spectral Flats

To investigate how well the calibration unit fulfills the goals for producing an adequate flat–field in the dispersion direction, the wavelength and temporal dependence of a number of flat–fields are considered.

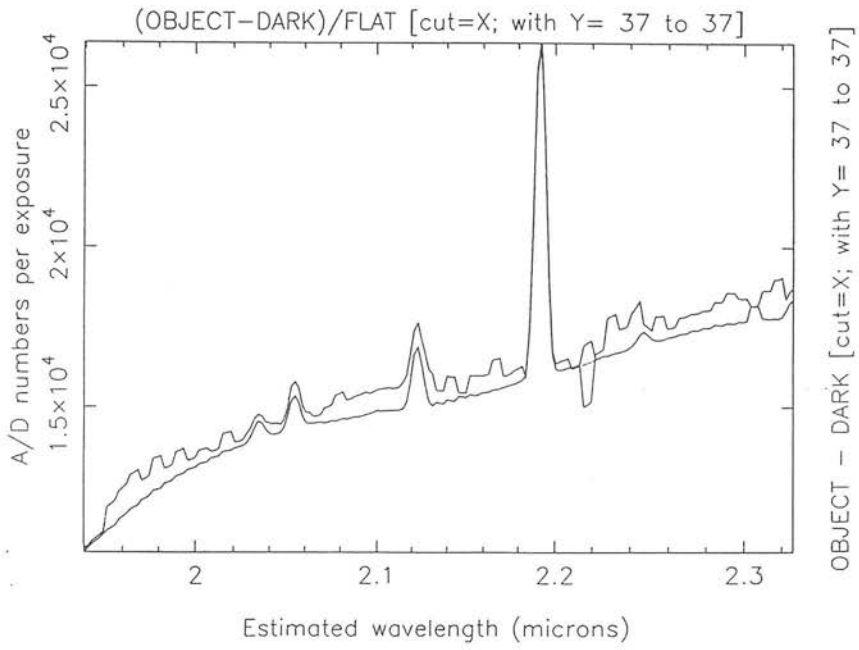


Figure 2.22: Two spectra of the krypton arc lamp, showing the difference made by dividing by a flat–field frame.

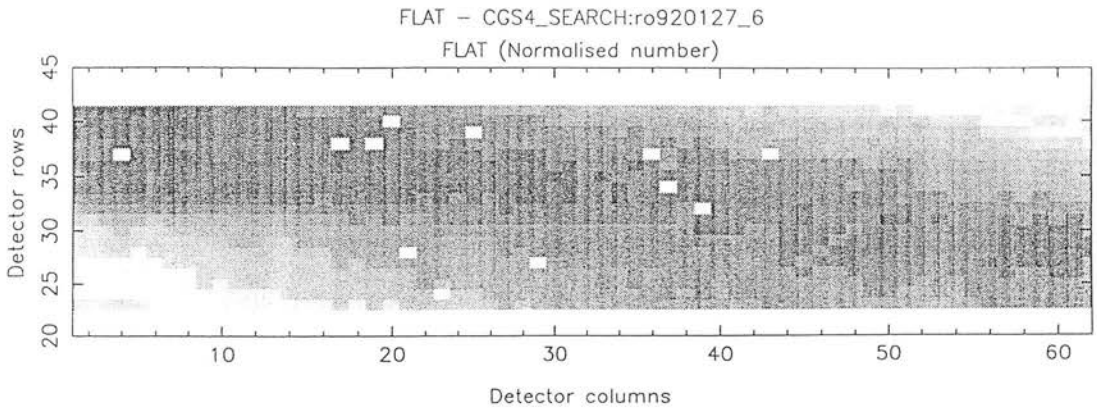


Figure 2.23: A typical, reduced flat–field frame. The 1σ variation in the pixel–pixel gains is $\sim$10%.

Figure 2.23 shows a typical flat–field frame taken with the low–resolution grating. From the histogram showing the distribution of the pixel gains it can be seen that the 1σ value for the change in quantum efficiency between the pixels is $\sim$10%. The accuracy with which an observation can be flat–fielded will be determined by both the signal/noise achievable with the black–body source, which depends on the temperature stability of the source, the photon noise on the black–body signal and the drift in the relative pixel gains during the observations.

The temporal stability of the flats (*i.e.* of the pixel gains and the black–body) can be demonstrated by dividing two flats at the same wavelength taken several hours apart. The frame shown in Figure 2.24 shows the result of dividing two flats taken on 28 February 1991, separated in time by $7^{h}00^{m}30^{s}$. The resulting gaussian distribution of pixel values has a mean of 1.000034 and a standard deviation of 7.6×10^{-3}, giving a 1σ drift in gain of less than 1% over a 7^{h} period. A similar comparison with a flat at this wavelength from the night of 1st March 1991, shows a spread in the ratio of the gains with a 1σ value of $\sim$2% (Figure 2.25). The short term variations in the flat–fields are demonstrated by the plot in Figure 2.26, which shows that adjacent flat observations

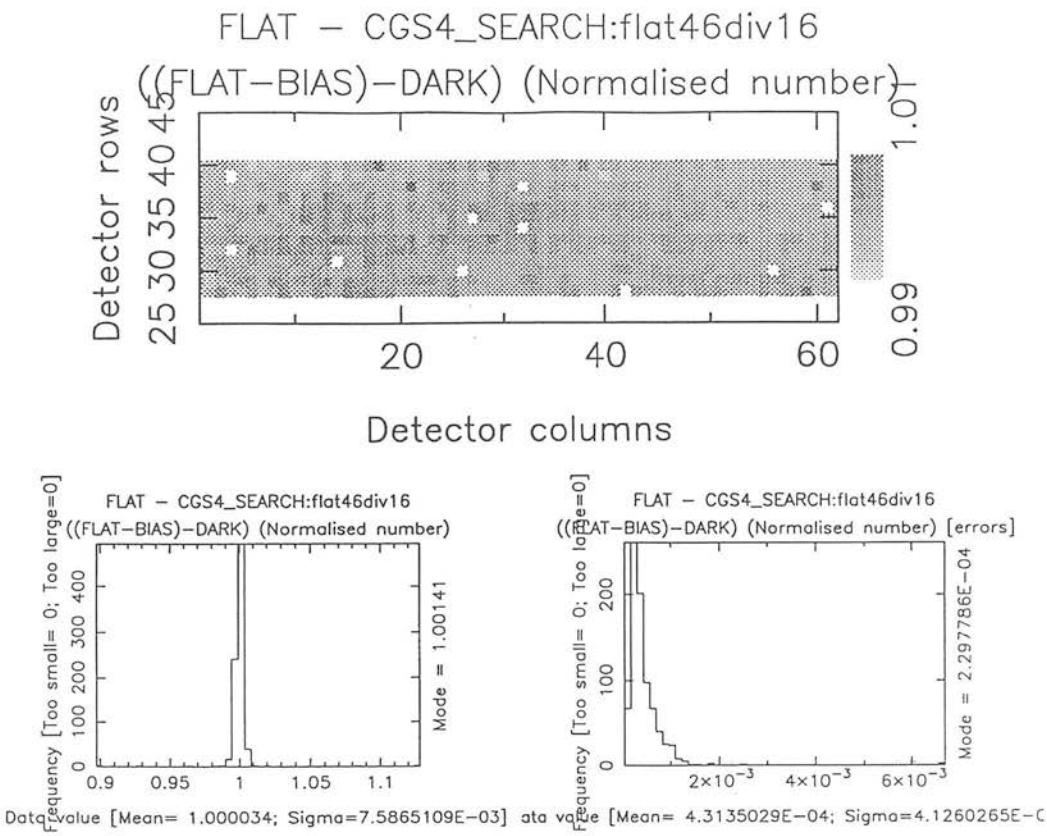


Figure 2.24: The ratio of two flat fields of the same wavelength observed 7hours apart. The 1σ drift in the relative values of the gains is less than 1%.

at the same wavelength show a 1σ change in the ratio of the gains of less than 0.2%. This was also true of the first flat and the third and fourth flats, so that the change in flats taken 13mins part was also less than 0.2%. These are all flat fields taken using the echelle.

The wavelength dependence of the flat–fields must be considered separately for echelle configurations and low–dispersion grating configurations because of the use of the CVF with the echelle. The CVF transmission at a given wavelength is constant along a radius, and the slit is aligned radially when there is no slit offset. If the slit is rotated, it is no longer aligned with a radius of constant CVF transmission, and the illumination of the array is altered. The frame in Figure 2.27 shows the result of ratioing two flat field frames of different wavelengths on the echelle: the first frame has wavelength 1.28410μm , the second 1.28839μm . Despite the very small change in wavelength, the ratio of the

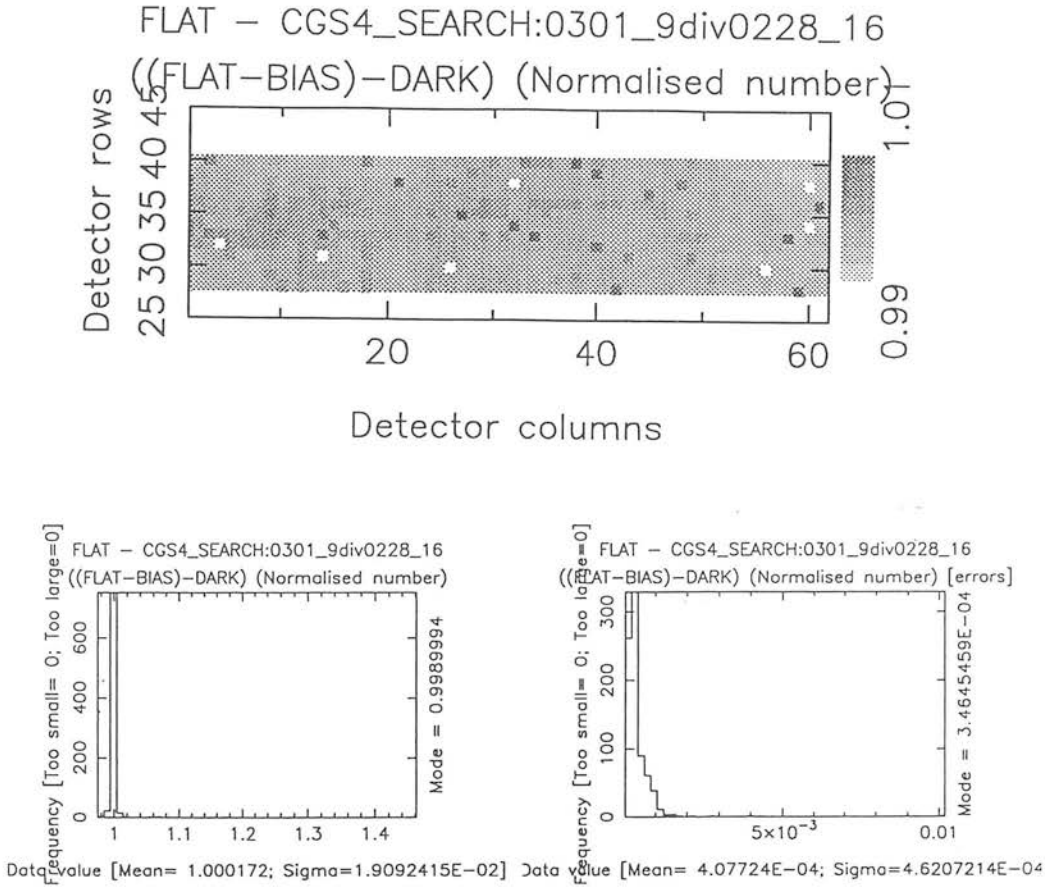


Figure 2.25: The ratio of two flat fields of the same wavelength taken on two consecutive days. The 1σ change in the gains is 2% in this case.

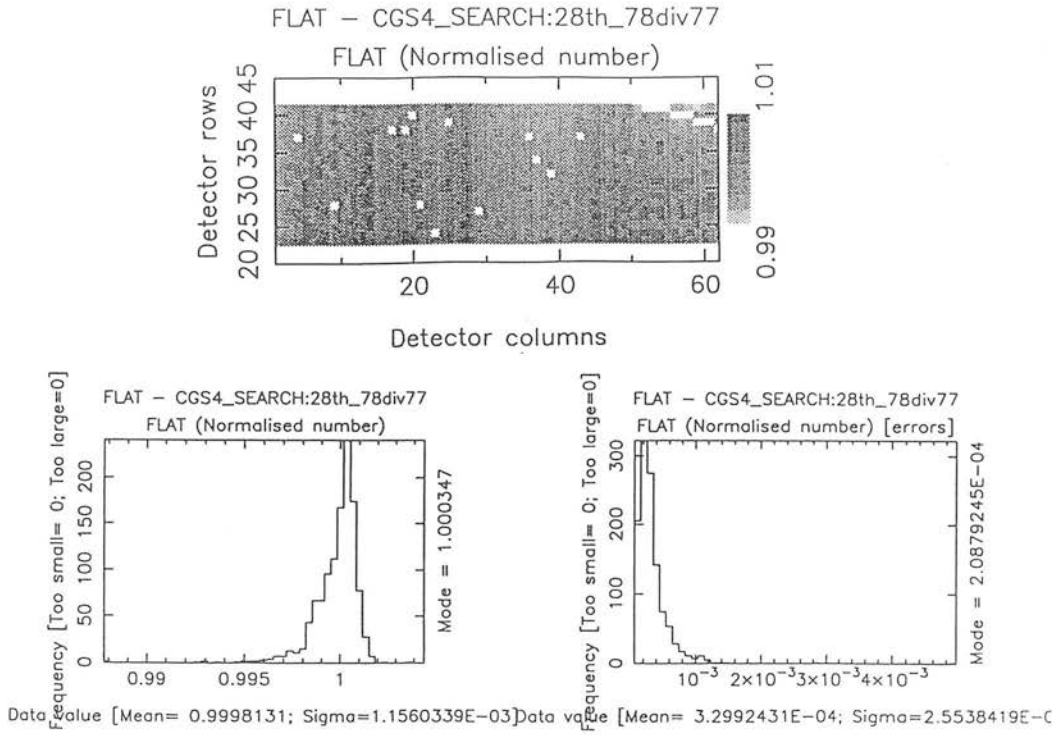


Figure 2.26: The ratio of two flats observed consecutively at the same wavelength on the echelle. The short term variations are ~0.2%.

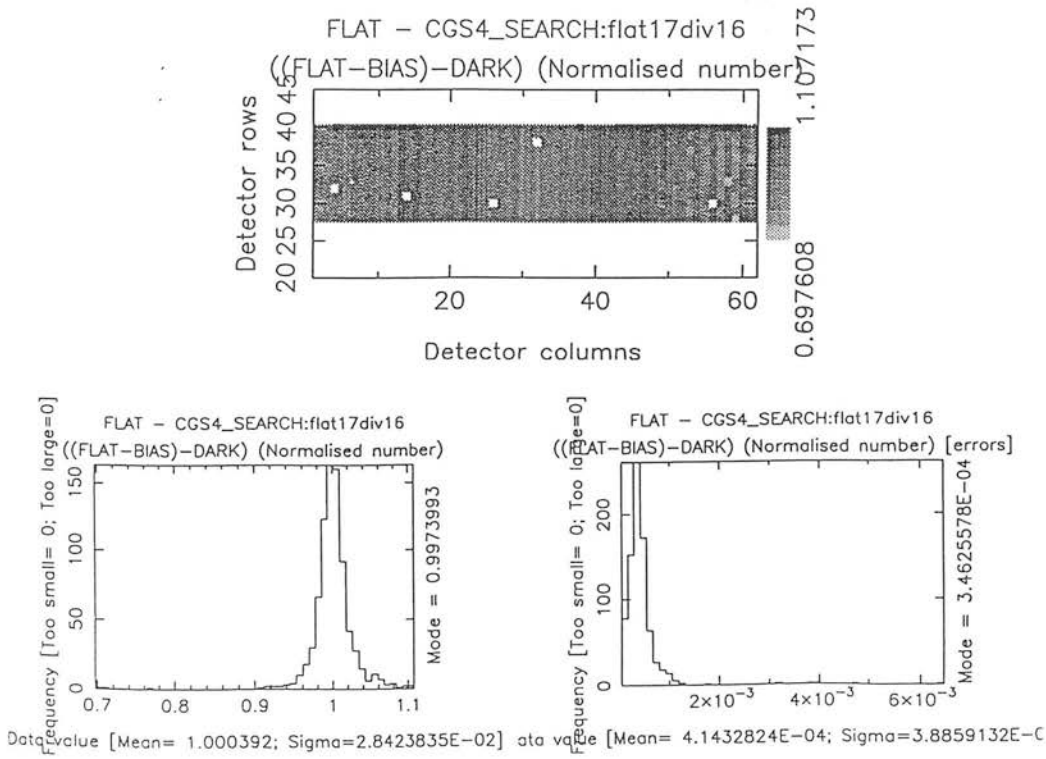


Figure 2.27: The ratio of two flat fields of the different wavelengths observed consecutively shows the variations of flat–field with wavelength for the echelle. The 1σ change in the gain is $\sim$3%.

flats shows some structure; the spread of values about a mean value of 1.000392 indicates that the 1σ ratio of the gains is less than 3%, although a difference of $\sim$10% can be seen for some pixels. Once again, this value remains constant with a flat taken 7 hours later and one taken the following night.

The variation in low dispersion flats with wavelength is seen by comparing three flats from the night of 23 February 1991 at 3.42μm , 1.60μm and 2.15μm . These flats are separated in time by hours, which, it is assumed, introduces a change in the relative gains comparable to that introduced in the echelle flats ($\sim$2%). The comparison between these different wavelengths shows that the 1σ change in gains can be as much as 20%. Thus, the wavelength dependence of the low–dispersion flats is very marked.

The signal/noise achievable on a flat–field frame is also significant when determining the accuracy to which the pixel–pixel gains can be removed from the object observation. Ideally, the noise on the flat should be low enough that it is dominated by the other noise sources for the observation; this can lead to a requirement that the flat be accurate to 1:10^4 for observations of the faintest objects. The flat with the greatest accuracy measured during tests is shown in Figure 2.28: it was observed on 13 February 1991 using NDR mode, with an on–chip exposure time of 1.006s and a total time for the observation of 201.2s. A dark frame of the same exposure time, also observed in NDR, was subtracted from it. The signal/noise measurement, taken as the ratio of the modal value for the data to the modal value for the errors, is 38652.5, giving a flat–field accuracy of 1:3.9×10^4.

The results of the above test can be summarised as follows. To obtain an accurate flat–field of the array for removing the relative pixel–pixel gains from astronomical observations, it is necessary to observe a flat–field for every new wavelength configuration. It is also advisable to observe a flat at the beginning and end of a set of astronomical observations, as the accuracy achieved in practice will be determined by the drift in the

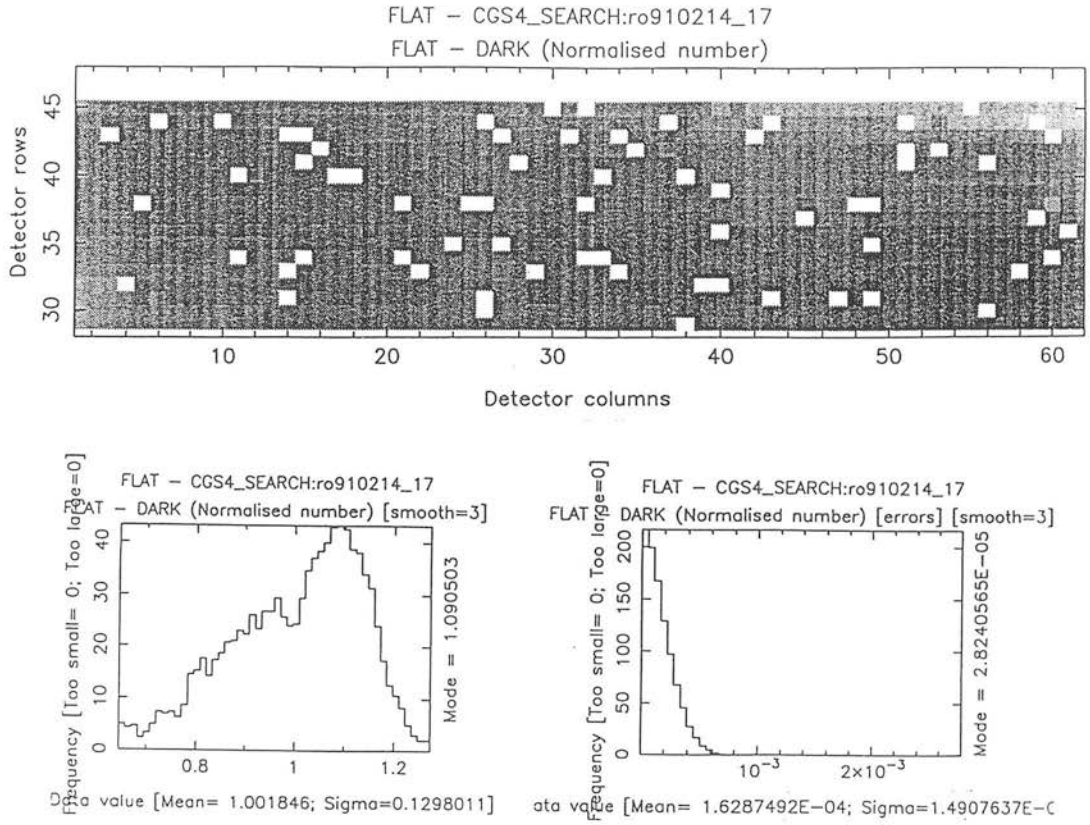


Figure 2.28: This flat–field frame shows the highest signal/noise of those measured during tests on CGS4. It was observed in NDR mode, with an on–chip exposure time of 1s and a total exposure time of 200s. The resulting errors are 1:3.9×10^4 of the black–body signal.

pixel gains with time (~0.1% per night).

2.12.3 Pixel–Pixel Gain Variations in the Cross-dispersion Direction

In order to verify that the output of the integrating sphere is flat along the slit and that the correct calibration of one pixel to another along the columns is achieved by dividing the observations by the function FF1(i,j) as defined in Section 2.12.1, the field from the CU must be compared with an alternative flat field. This allows an independent determination of the vignetting function, V_j.

The search for a second uniform source, highly uniform along the slit, but not necessarily spectrally as it may be combined with the spectral flat, lead to consideration of

using the sky in the spatial direction thus turning the IR background to some advantage. Moreover, since for every observation of the source there is a corresponding sky observation made, no further observations are required if there is sufficient signal/noise on the sky frames (1:10^4). It is possible to make a first order estimate of the spatial flatness of the background emission. The two distinct regimes should be considered separately. In the wavelength region where background emission is dominated by the non–thermal emission from OH (1–2.3μm), the spatial intensity profile is determined by the nature of gravity waves which propagate through the emission layer (80–100km) exciting the OH. (This phenomenon and its effect on the OH emission strength is discussed more fully in Chapter 4.) This activity results in the formation of bright, aurorae–like emission structures imposed on the night sky, whose spatial wavelength is ~20–30km (see *e.g.* Taylor *et al.* (1987) for imaging and analysis of these waves). Assuming a model for this emission structure as a sine wave of wavelength 25km at a height of 85km running perpendicular to the slit with an intensity variation of 10%, consistent with theory and observation (Chapter 4, Ramsay, Mountain & Geballe 1992), gives an intensity change of ~1:10^4 in the cross–dispersion direction, which is sufficient for most observations. However, the presence of high–altitude winds of even a few ms^{-1} would reduce the stability of the structure, as the wave pattern would cross the beam in a few seconds. Each night should be considered as a separate case until a more comprehensive study can be made of the OH emission from Mauna Kea. A limit to this technique is likely will be the lack of signal for some observing configurations. For example, observations with the echelle may have no sky lines within the wavelength range observed. Even when OH lines are observed the observations may be detector noise limited.

For the thermal infrared (2.3–5.0μm), the sky background is dominated by the rapidly changing thermal emission from the sky. As for the telescope dome, changes in sky temperature of ~10^{-3}K will reduce the accuracy for the flat–field to below the estimated accuracy required (1:10^4), making this flat–fielding technique unsuitable for thermal wavelengths.

2.12.4 Method

In order to achieve the correct calibration of one pixel to another along the slit, by determining the vignetting function $V(j)$, essentially the same process as described in Section 2.12.1 can be performed but averaging over columns (i) instead of rows to get the profile $lamp\_col\_avg(j)$:

$$lamp\_col\_avg(j) = I_{bb}(j) * V(j) * < W(i) * B_{\lambda}(i) * G(i,j) >_i . \tag{2.44}$$

Again we can fit a smooth polynomial and evaluate this at each pixel position to obtain:

$$lamp\_col\_pol(j) = I_{bb}(j) * V(j) * < W(i) * B_{\lambda}(i) >_i . \tag{2.45}$$

where, as above, the polynomial smooths over the variations in <... G(i,j)>.

If the output of the lamp/integrating sphere did provide flat illumination along the slit, I_{bb} would be a constant and the vignetting function V(j) would be determined with arbitrary normalization. Then division of the columns of FF1(i,j) by $lamp\_col\_pol(j)$ would produce the flat-field gain matrix, G(i,j). However, variations in the lamp intensity along the slit of at least 2% have been seen which imply that I_{bb} is a slowly varying function of row j and hence V(j) must be determined independently; this requires the sky observation.

Taking the sky observation sky(i,j), we can determine the vignetting V(j) by summing along columns i as before to form an average:

$$sky\_col\_avg(j) = V(j) * < W(i) * sky_{\lambda}(i) * G(i,j) >_i . \tag{2.46}$$

Fitting a smooth polynomial $sky\_col\_pol(j)$ gives V(j) with an arbitrary normalization. This function can used to correct the partial flat field FF1(i,j) by dividing by the column-averaged lamp signal and multiplying by the vignetting function to get:

$$FF(i,j) = FF1(i,j) * sky\_col\_pol(j) \div lamp\_col\_pol(j) = G(i,j) * V(j). \tag{2.47}$$

This provides the gain matrix G multiplied by the spatial vignetting function V, which is sufficient as all the astronomical observations made are of objects outside the system. The detected signal from an astronomical source of intensity I_S(i,j) is:

$$obs(i,j) = \{I_S(i,j) * T(i) + sky(i)\} * V(j) * W(i) * G(i,j) \tag{2.48}$$

where sky(i) is the spectrum of the night sky plus the thermal emission from the surroundings outside CGS4 and T(i) is the atmospheric transmission. After division of this signal by the flat-field FF(i,j) we get the flat-fielded result:

$$obsFF(i,j) = \{I_S(i,j) * T(i) + sky(i)\} * W(i). \tag{2.49}$$

Thus we have removed the effects of filters, vignetting and the different gains of the pixels. To get to the intrinsic properties of the source I_S itself, however, we need to observe a standard star at similar airmass.

2.12.5 Calibration: observations of a standard star

To remove the effects of telluric absorption we divide the object spectrum by that of a standard star, whose intrinsic spectrum is assumed to be well understood (usually approximating to a black-body). The flat-fielded signal from such a star of intrinsic intensity I_* can be written as:

$$starFF(i,j) = \{I_*(i,j) * T(i) + sky(i)\} * W(i). \tag{2.50}$$

The star signal may be 2-dimensional because of the finite size of the seeing disk, therefore the star spectrum is extracted by averaging rows, either normally or by optimal extraction (Horne 1986), and subtracting the background to obtain the 1-D spectrum:

$$starFF\_sub(i) = I_*(i) * T(i) * W(i). \tag{2.51}$$

The flat-fielded, sky–subtracted object observation is:

$$obsFF\_sub(i,j) = I_S(i,j) * T(i) * W(i). \tag{2.52}$$

Dividing each row j of obsFF_SUB by starFF_sub yields the fully calibrated spectrum:

$$obsFF\_sub\_div(i,j) = I_S(i,j) \div I_*(i) \tag{2.53}$$

2.12.6 Spatial Flats

An estimate of the stability of the spatial distribution of the emission from the sky was given in Section 2.12.3. The procedure adopted to determine the spatial flat field in practice was as follows. A group of H–band observations with the 75 l/mm grating was chosen, as the window is filled with bright OH emission lines and signal/noise on the sky spectra is not, therefore, a limiting factor. Each column for each of the 10 sky observations was summed and a polynomial fit to each spectrum was found. A linear fit was seen to be appropriate. The polynomial which describes the spatial shape when observing the calibration unit and continuum sources ($lamp\_col\_pol(j)$) is obtained by summing along all the columns of the flat. The spatial profiles found for the sky and the black–body in this way are shown in Figure ??; both the fits are linear. The gradient of the black–body profile along the slit is 0.03±0.01 and that for the sky profile is -0.08±0.01. The sky profile is subject to the telescope vignetting, the instrument vignetting and may also show a real gradient of pixels gains along the slit. The black–body profile is not subject to the telescope vignetting, but additionally is subject to the integrating sphere profile. A combination of both profiles is then required to retrieve the pixel–pixel gains. The ratio of the two profiles gives a gradient of 2.7 with a 1σ error, hence the change in gradient for the two profiles is barely significant at the 3σ level, implying that the black–body plus integrating sphere gives a good approximation to the pixel gains along the slit.

According to the formula in Section 2.12.4, the original flat–field is then divided by $lamp\_col\_pol(j)$ and multiplied by $sky\_col\_pol(j)$, which are normalised to unity. The resulting flat field is plotted in Figure 2.30. Observations which are reduced automatically

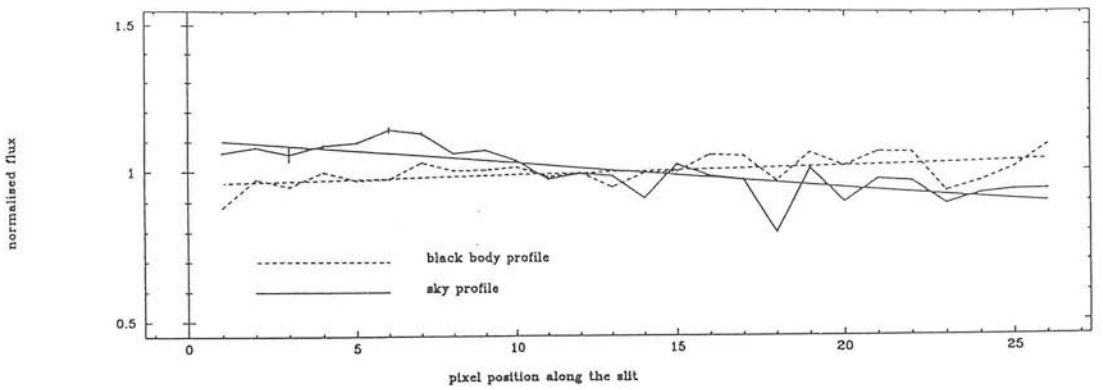


Figure 2.29: The spatial profiles for observations of the sky and the black–body source and intergrating sphere. The gradient from the integrating sphere is 0.03±0.01, from the sky it is -0.08±0.01.

by the CGS4 reduction software can be fully flat fielded by multiplying either the reduced group or a single reduced observation by $lamp\_col\_pol(j)$ and dividing by $sky\_col\_pol(j)$.

2.13 Conclusions

The details of the optical design of CGS4 have been presented, along with the motivation for that design. Aspects of noise sources for NIR spectroscopy have been discussed, including the detector dark current and dewar background, the read–out noise of the detector and the sky background emissions. Cooling CGS4 to LN_2 temperatures, and including the motors which control the optical components within the cryostat has lead to a dark current of $\sim 1e^{-}s^{-1}$. A reduction in read–noise has been achieved by sampling the signal on the array multiple times before reading out the charge. For a 35s exposure, the effective read–out noise is $34e^{-}$.

With the knowledge of these noise contributions, a model for the sensitivities of CGS4 was constructed, using model values for the sky background and estimates of the instrument throughput. Measured sensitivities show that, at thermal wavelengths the

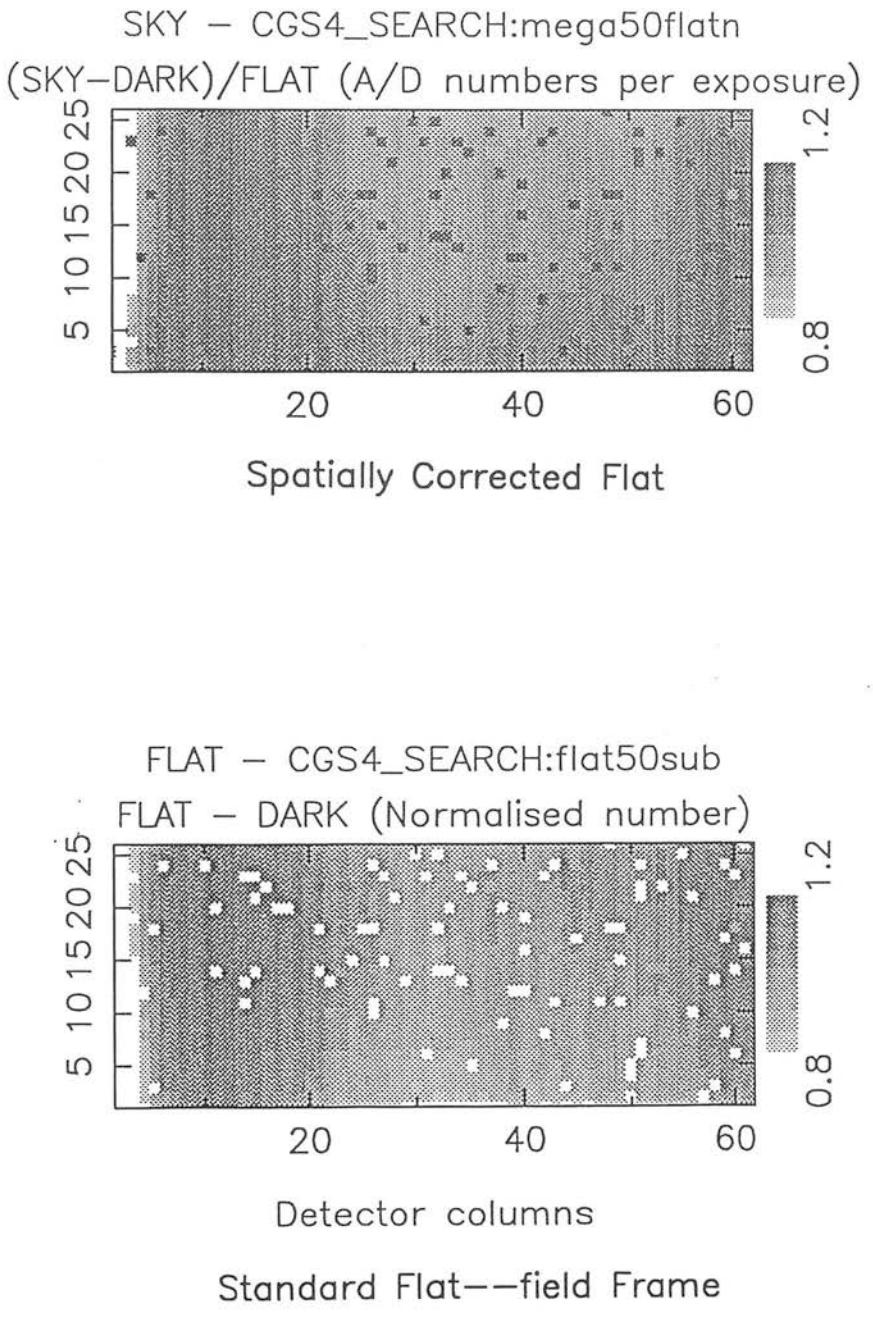


Figure 2.30: A comparison of the partial flat–field frame, created from an observation of the black–body source in the CU, and the flat field created by combining the vignetting profile derived from observations of the sky with the partial flat.

sensitivity for a 3σ30min observation agrees well with the estimated values. At L, the difference in magnitude was 0.3; at M it was 0.7. At K, a larger discrepancy was observed (1.4 magnitudes), probabaly due to changing sky background levels due to variations in the OH emission.

To flat–field the CGS4 array both in the dispersion and cross–dispersion directions, a calibration unit was designed. The $90\widehat{\text{m}}$ field is illuminated uniformly by the output from an integrating sphere, re–imaged onto the slit–plane. The continuum sources in the calibration unit were chosen to provide high signal/noise measurements, as flat–fielding accuracies of $1{:}10^4$ may be required. Tests carried out with the instrument show that a flat–field frame with errors of $1{:}4{\times}10^4$ can be achieved. Temporal Drifts in the pixel gains and black–body signal of $\sim$0.1% require that flat–fielding be carried out both before and after the astronomical observations to be most effective. The success of the CU at providing a spatial distribution of illumination on the array the same as that from astronomical observations has been proven. Thus, the flat–field corrects for pixel–pixel gain variations both spatially and spectrally.

Chapter 3

The Testing and Calibration of CGS4

Introduction

This Chapter describes the alignment and calibration of Cooled Grating Spectrometer 4, carried out at the Royal Observatory in Edinburgh in the period from October 1989 to December 1990 and thereafter at the Joint Astronomy Centre, Hawaii, before CGS4 was commissioned on UKIRT. These tests fall into two categories; "warm tests" consisting of the alignment of the optical components and tests of the image quality using laser light, and "cold tests" carried out on the cooled cryostat. The former consist of "static" tests of the optical components: determination of the dimensions of each component, their assembly into modules and the internal alignment of each module, and finally the integration of all the modules into the complete instrument (Figure 3.1). The latter comprised calibration of the instrument using infrared sources, examination of the image quality and tests of the reliability and repeatability of the calibration. As the motor control, array control, data acquisition and reduction software were used exclusively to drive the instrument, these too were thoroughly exercised. The aim in adopting a mechanical approach to the alignment of CGS4 was to provide a reliable method of aligning the instrument for maximum throughput without vignetting and excellent

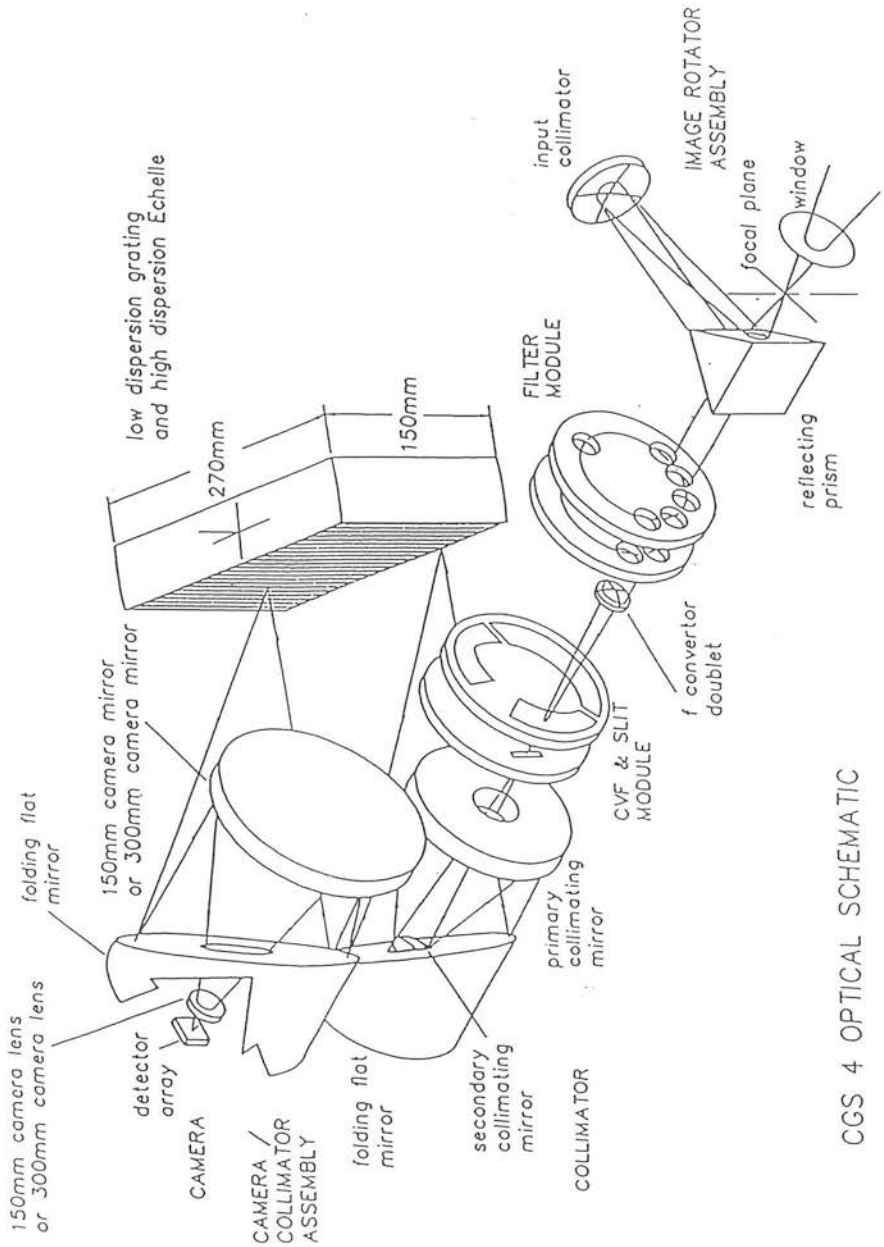


Figure 3.1: The optical path of CGS4, showing the location of the modules and the optical components, the alignment of which is described in this Chapter.

image quality, to ensure that the alignment would be unaffected by cooling the optical components to 70K, and to allow the modules of the instrument to be interchanged freely with confidence that no lengthy realignment would be required. The philosophy behind the calibration and its interface with the software used to control the instrument configuration was that, as far as possible, no knowledge of the details of the instrument would be required to observe with CGS4. In practice this means that only the wavelength of observation and the position angle of the astronomical object should be entered by the observer, thereafter the entire motor configuration should be derived in software. The exposure times and other parameters pertaining to the detector read–out must also be entered by the astronomer.

The first test of the optical components, not described in this thesis, was an inspection of their dimensions, parallelism *e t c* by Ferranti, and later at ROE, using a three axis measuring machine, which confirmed that all the machined components complied with the specifications before testing began. Good use can be made of the fact that, for example, the front and rear faces are known to be parallel, allowing measurements to be made with reference to the most accessible datum surface where space and access were limited.

Many techniques and experimental methods are common to the alignment of each component; these are detailed in Section 3.1 along with a description of the principal pieces of equipment used. Subsequent Sections outline the alignment of the individual modules and their integration into a complete optical system. The results of testing the optical alignment after the first period that the instrument was cooled (the first "cool–down") are given in Section 3.8. The warm tests and calibration of the instrument are described fully in Section 3.9.

3.1 General Techniques and Equipment for Alignment

The global method behind the mechanical alignment is to determine if the alignment pads which locate the modules on the datum surfaces are correctly sized, and to alter them appropriately if needed. To this end, the decentre and tilt of a given module

relative to the optic axis must be measured. The three main resources for the optical alignment are the HeNe laser, the Rank Taylor Hobson Alignment Telescope (AT) and the f/36 simulator.

3.1.1 The HeNe laser

A Melles Griot HeNe 5mW laser (λ=632.8nm) was used for all the warm tests, when IR light cannot be used. Since the CGS4 optics are all reflective, with the exception of the ZnSe field lens in the short focal length camera and the BaF_2/CaF_2 f/10 convertor doublet, optical light can effectively be used to measure the linear and angular alignment of the optical components. The refractive index of the f/10 convertor lens is almost constant between at 632.8nm compared with the value for 1–5μm , but the ZnSe refractive index changes significantly. It is possible to confirm that the camera focal length is as predicted for optical light. The spot sizes seen with the laser will be larger than the 30μm spot sizes required in the NIR (Section 2.2), however, measuring the image size and the change in size across the array remains a useful test of the alignment and verifies that no gross errors in the optics are present. The final test of the focal lengths and image quality can only be carried out on the cold instrument.

The laser was used in conjunction with a beam expander to give a diverging beam and a beam splitter for measuring foci accurately. The beam splitter or pellicle was used with a pinhole mounted in front of the laser output; an image of the pinhole was formed at a screen on the pellicle holder. A return image of the laser spot was coincident with this image if a plane mirror was placed at the exact focus of the beam. One example of a test using this arrangement is the measurement of the camera back focal length (Section 3.5.1).

3.1.2 The Alignment Telescope

To quantify the tilt and decentre of the modules to within the specified tolerances (Table 3.1) requires the accuracy achievable with an alignment telescope. The RTH alignment telescope will measure positions to within 10μm over distances of $\sim$1m. The

decentre of the AT can been measured using two micrometer scales which control motion in the x and y directions, referred to in the tests that follow as Dial A and Dial B; the tilt is controlled by two further micrometers on the AT mount which do not alter the decentre. The AT has three modes of operation: auto–reflect, auto–collimation and alignment mode. The tilt can be measured in auto–reflect mode using the illuminated coverglass target which shows a series of concentric rings. When a mirror is placed in the beam, a return image of the crosswires is seen on the coverglass. The radial distance of the return image from the centre can be read from the concentric rings, which are calibrated and labelled with their radius from the centre in mm (Figure 3.2). The tilt (mrad) is then:

$$\text{tilt} = \frac{\text{circle value}}{\text{baseline}} \tag{3.1}$$

where the baseline is the distance between the AT coverglass and the mirror. Auto–collimation works in the same way, but the AT is focused at infinity so that the crosswire is reflected from the target and re–imaged at the AT. It is not possible to quantify the tilt detected in this way, but it can be used as a quick and accurate test of squareness. In alignment mode, the AT is focussed on a target at the module and the displacement between them measured by adjusting the barrel positions of Dials A and B until the crosswires overlap.

3.1.3 Use of the f/36 Simulator

To simulate the field from UKIRT and the alignment of the instrument to the telescope, a device referred to as the f/36 simulator was used. The simulator produces an image at the Lyot stop which fills the field as would the image of the telescope secondary at a distance of 11m, and which comes to a focus at the slit plane. This device consists of a small flat mounted behind an aperture stop and a spherical mirror, radius of curvature 1000mm (Figure 3.3). At the input a pinhole is illuminated by the laser; the output is an f/36 beam of focal length 1000mm at a height of 100mm above the optical bench, the designed height of the optic axis at the input window of CGS4. The illumination of the f/36 simulator entrance pinhole by the laser is adjusted using the micrometers on the

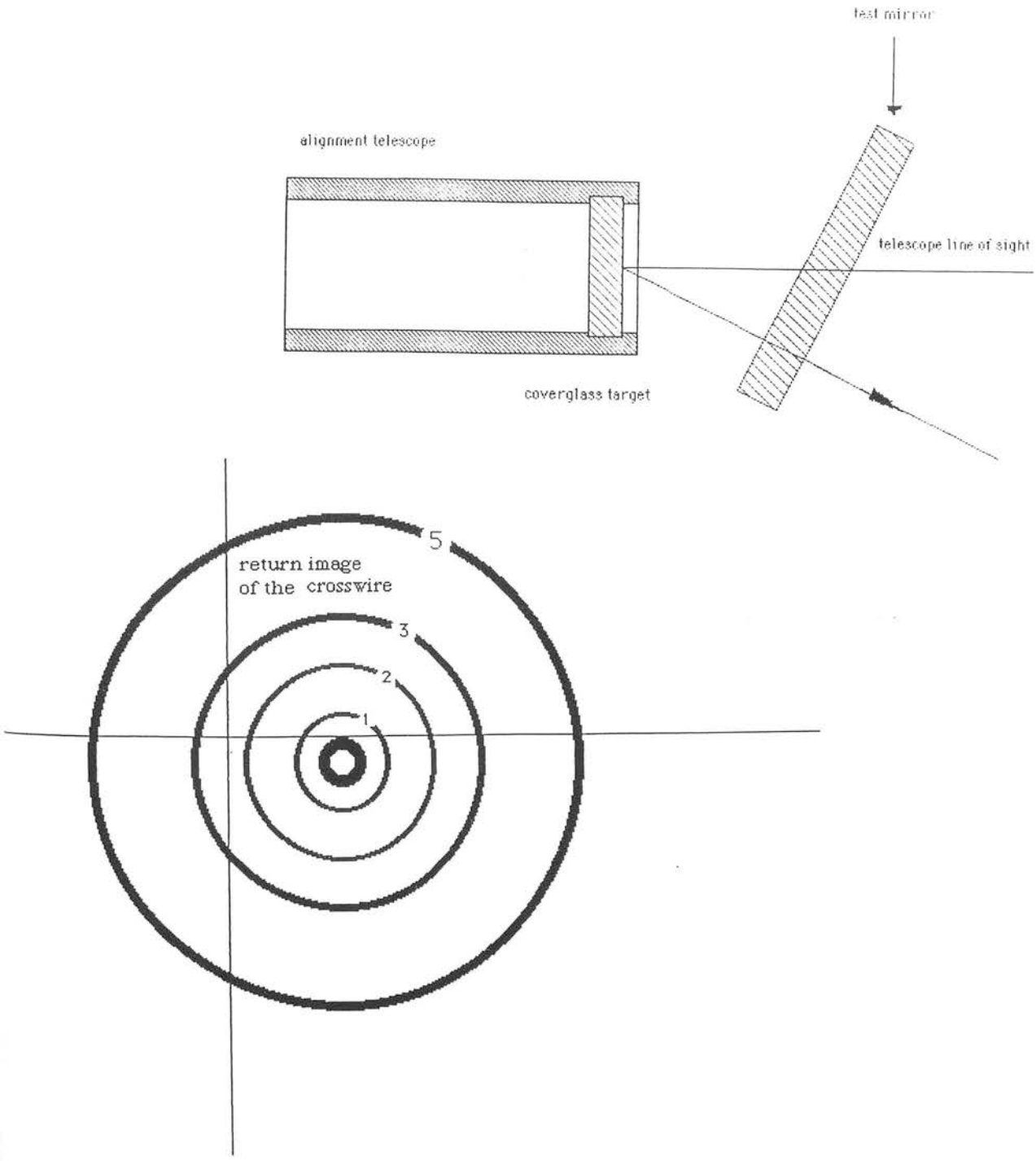


Figure 3.2: The principle of the operation of the alignment telescope in auto–reflect mode. With the test mirror tilted with respect to the AT line of sight, as shown, the return image of the crosswire appears displace from the centre of the target by some radial distance which may be determined using the calibrated concentric rings

	Power (Fringes)	Irregularity (Fringes)	Clear Aperture$^{\phi}$ (mm)	Tilt$^{\pm}$ (mrad)	DEC$^{\pm}$ (mm)
1st flat	1	1/2	91.8	0.27	0.1
Input Collimator	1	3/4	69.4	-	0.05
2nd flat	1	1/2	34.3	0.27	0.1
F/converter lens	4	1	30	1	0.1
Secondary Collimator	2	3/4	28.4	0.49	0.045
Primary Collimator	2	3/4	134.1	0.27	0.025
1st compound flat	1	1/2	142.6	0.34	0.025
Grating	3	1/2	143x258.7	0.72	0.1
2nd compound flat	1	1/2	~250	0.34	0.025
Camera Mirror	4	1/2	~250	0.27	0.035
Camera lens:					
1st surface	1	1/4	39.2	0.41	0.053
2nd surface	1	1/2	39.2		

Table 3.1: The tolerances on aligning the optical components of CGS4, derived from the requirement that the spot size be $\sim 30\mu$m .

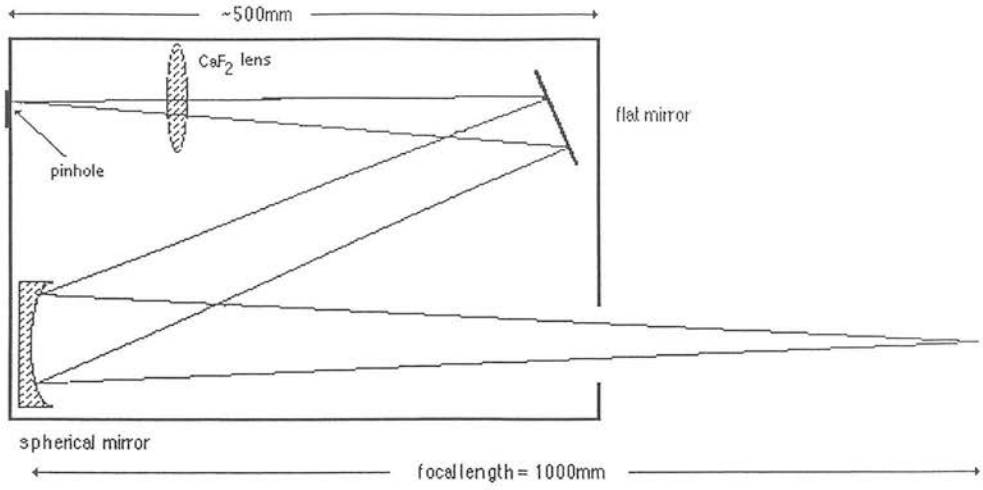


Figure 3.3: A schematic diagram of the f/36 simulator, which fills the $90''$ field of CGS4 in the same way that UKIRT does.

laser mount until it appears to be symmetrically illuminated and the small angled flat, which acts as an aperture stop, is lit with no diffraction rings in evidence. The beam can be positioned to evenly fill the stop, using the stop micrometers, until an image appears centrally placed on the concave mirror. A CaF_2 lens is placed in the beam before the stop to correct for the astigmatism inherent in the source. The exact position of focus relative to the output face of the f/36 simulator must be found empirically; it was found to be 486±3mm from the face of the f/36 simulator. The f/36 simulator was often used with a dichroic mirror, which simulated the dichroic mirror on UKIRT, placed in the beam. The dichroic mirror micrometers were then used to move the beam across the field defined by the input mirror of CGS4.

3.2 General Procedure

A procedure repeated many times during the optical testing was the alignment of the AT and/or f/36 simulator to the optical components being tested; the method is outlined here to prevent repetition. Before beginning the optical tests it must be ensured that the test equipment is set–up to be perpendicular to, and centred on, the optic axis. The alignment telescope can be used in both auto–reflect and alignment mode to achieve this, as described in Section 3.1.2. Using a perspex target mounted on the component under study, with the AT in alignment mode, the xy positions are adjusted using Dials A and B until the crosswires are centred on the target pinhole or crosswire. Changing the AT to auto–reflect mode and replacing the target with a flat mirror, the return image of the crosswires will lie at some angular distance from the centre of the coverglass and can be aligned using the tilt adjustment barrels on the AT mount. This does not alter the linear adjustment. The procedure quickly iterates to a set–up where both the angular and linear adjustments are correct; the component is then aligned and the datum barrel positions for Dials A and B can be noted.

The f/36 simulator is aligned to the optics using a double pass of the beam through the simulator. Using first a target on the optical component, the position of the laser spot should be centred by moving the the dichroic mirror. The tilt is tested by replacing the target with a mirror and checking the return path of the beam through the simulator and corrected using the micrometer controls on the dichroic.

In both of these cases the mirrors used are optical quality flats (accurate to $\sim \frac{\lambda}{100}$) and the targets are perspex. The target centres are marked with crosswires or a pinhole which then define the optic axis. To eliminate any errors due to misfitting or eccentric targets or rough patches on the mirrors which would give a tilt error, both the targets and mirror are rotated, the measurements repeated and the average of the two values used. This precaution is also used when measuring the datums.

Other procedures in common to all measurements are those intended to reduce errors. The tolerances on tilt and linear displacement are so severe that any grit, shard of metal or raised area on the optical reference surfaces could introduce significant misalignment.

For example, if the surface is raised by only 5μm this will appear as an angular misalignment of 5mrad measured over 1m, which is comparable to the typical tolerances of 25mrad. Before any measurement, datum surfaces were cleaned and checked for burs by drawing a straight–edge along them. The equipment and optics were clamped firmly to the bench, so that they were not disturbed during the tests. In many of the measurements described below, errors are reduced by using a long baseline between the AT and the optics, so full use of the length of the optical bench was made. One of the desirable features of this mechanical alignment is that the modules should be able to be exchanged with the confidence that the alignment has not altered. Thus the final test of the alignment was a check on repeatability; the module was removed and then replaced and the measurements repeated.

3.3 The Image Rotator Alignment

The image rotator (IROT) supplies a 2α rotation of the image of the slit on the sky for an α rotation of the image rotator by virtue of a three mirror arrangement (Figure 3.4). With the image rotator correcting aligned, the position of an image in the slit plane will be invariant on rotation of the IROT to within a linear tolerance of 0.25mm and an angular tolerance of 0.5mrad. Misalignment of the image rotator will cause the image of the slit on the sky to move relative to the object, resulting in loss of light from compact sources.

The IROT consists of two flat mirrors mounted back to back in a prism configuration on a base–plate and a third, collimating mirror mounted above them. The position of the collimating mirror must be adjusted until the emergent collimated beam lies within 0.25mm of the optic axis and suffers less than 0.5mrad tilt. This tolerance applies for all image rotator angles, thus rotating the IROT by 90° should not cause the tilt and decentre to move outside the tolerance. The complete test set–up is shown in Figure 3.5.

The IROT is attached to the casting by a bearing housing mounted on six M6 screws. During the tests the bearing housing was mounted on an angled plate which was then firmly clamped to the optical bench. The AT is set–up as outlined in Section 3.2 using

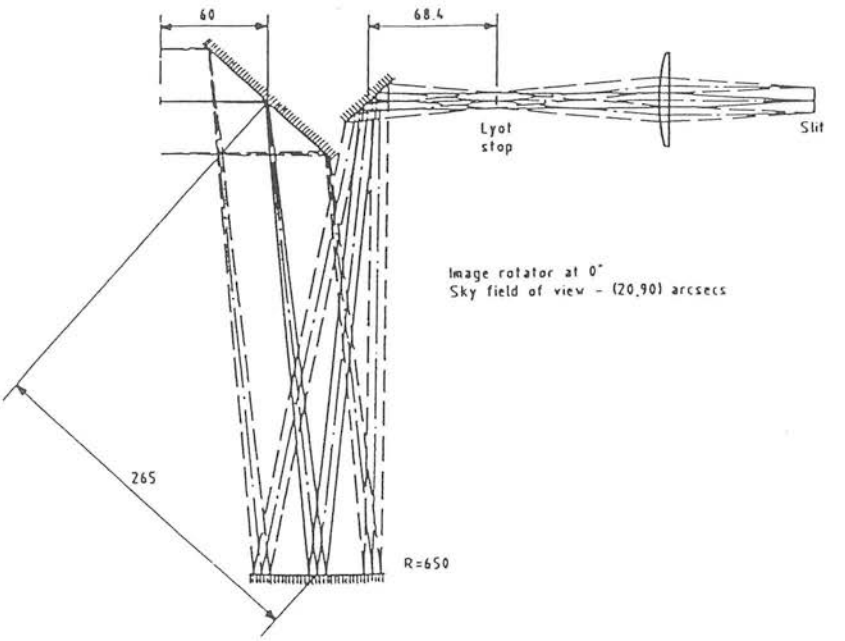
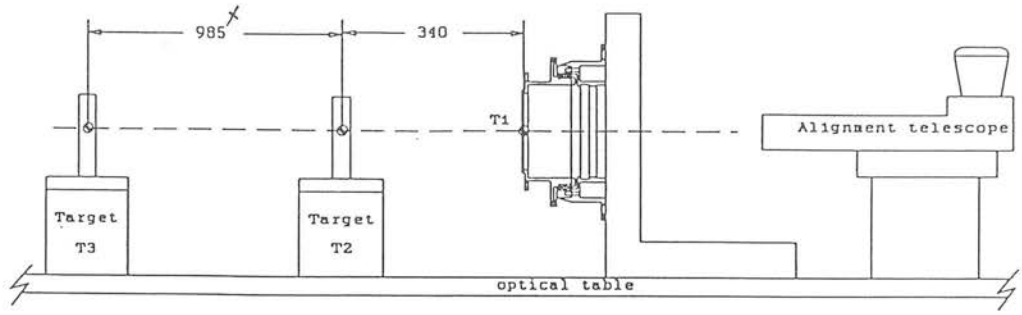


Figure 3.4: The optical layout of the image rotator, which rotates the image of the slit on the sky through 2α for an α rotation of the image rotator.



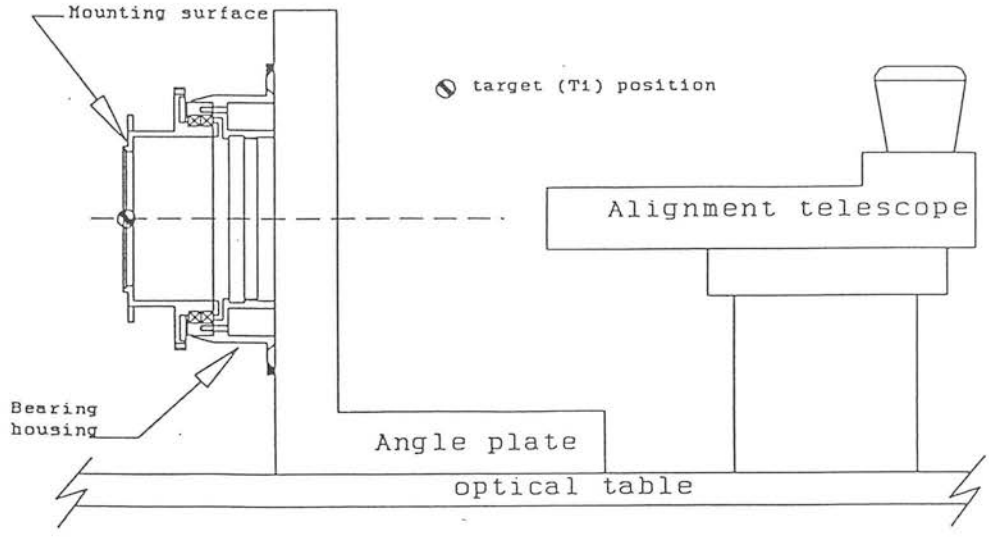


Figure 3.5: The test set–up for the image rotator alignment tests, showing the alignment of the AT to the bearing housing and the positions of the targets, T2 and T3, used to determine the tilt and decentre of the IROT.

a target (T1) mounted at the rear face of the housing and a flat mirror held against this surface but offset by two gauge bars. This defines the optic axis for the remainder of the experiment. Two further targets (T2, T3), mounted on XY adjustable stands a known distance apart, are then used and must be aligned to the defined optic axis using the XY micrometer adjustments. The AT is now known to be perpendicular to the targets which all lie on the optic axis. The micrometer readings on Dials A and B for T2 and T3 are noted as datums. The IROT can then be assembled and mounted on the bearing housing. With the image rotator in the beam the targets appear to be displaced from the optic axis; they are returned to it by adjusting the micrometers and the new positions noted. This is repeated at 90° rotation. The offset from datum of the new image position defines a linear error, Δz and can be used to determine an angular displacement, $\Delta\alpha$.

$$\Delta z = (T3_i - T3_{id}) - (T2_i - T2_{id}) \tag{3.2}$$

$$\Delta\alpha = \frac{\Delta z_i}{d} \tag{3.3}$$

for each combination, i of X,Y at 0° and 90°, where $T3_{id}$ and $T2_{id}$ are the datums for the targets, $T2_i$ and $T3_i$ are the measured positions for the targets with the IROT mounted, d is the distance between T2 and T3, which was 625mm when the measurements in Table 3.2 were made. Initial misalignment was corrected by adjusting the position of the collimating mirror in its mount, where it is located by three reference pads. The mirror was moved until the AT crosswire coincided with targets T2 and T3. Alignment was only finally achieved when 15μm was scraped from the front, leftmost pad and 12μm scrapped from the back rightmost pad of the mirror mount (looking from the input). The final measurements are shown in Table 3.2.

3.4 The Lyot Stop

The lyot stop is a cold baffle which acts as a field stop and serves to reject any stray light which has not been rejected at the input baffles. It is mounted in the input face

Image Rotator Alignment Results							
target	Datums	90°	0°	90°	0°	decentre Δz (mm)	tilt $\Delta\alpha$ (mrad)
T_{2x}	1.56	1.69	1.68	Δx	Δx	90°	
T_{2y}	4.13	3.88	4.23	0.14	0.13	0.15	0.24
T_{3x}	3.24	3.23	3.23	Δy	Δy	0°	
T_{3y}	6.64	6.34	6.55	0.05	0.19	0.23	0.37

Table 3.2: Results of the image rotator alignment tests, giving angular and linear errors within the acceptable tolerances (0.25mm and 0.5mrad). The "90°" columns indicate measurements taken with the image rotated 90° from horizontal; the horizontal measurements are given in the "0°" columns. The difference between the x/y offsets for the two targets is given by Δx/Δy The baseline for these measurements was 625mm.

of the filter module, at a theoretical distance of 68.4mm from the output mirror of the IROT (Figure 3.4). The beam passing through the lyot stop has been collimated by the IROT and should be 9mm in diameter (the physical size of the stop is 9.6mm). Both the degree of collimation and the shape of the beam must be checked at this position.

The output of the IROT at the lyot stop position can be examined without the stop or module in place. Qualitative checks of the beam shape and degree of collimation of the beam were made using a screen scribed with an 11mm diameter circle translated in the beam. A check was made for ellipticity of the beam; if the beam does not pass completely through the circular stop, the throughput will be reduced. For distances less than 1m from the IROT, the beam is consistently 9mm in diameter. This check was repeated for on and off–axis rays with the filter module mounted on the casting and a target of concentric circles placed in the lyot stop position. With the baffle tube inserted at the input window, the edges of the first baffle acted as a guide to the width of the beam; the laser spot is tracked from one edge to the other both horizontally and vertically. No ellipticity or vignetting was seen at any extreme.

3.5 Assembly of the Slit and Filter Modules

The assembly of the filter module is relatively straightforward. The two filter wheels contain broad and narrow band filters covering the 1–5μm range of the spectrometer, each of which has a 1" diameter and is assigned a location around the wheel. The filter is held in place by an anodised, sprung baffle with the coated side facing towards the slit wheel. There is one blank and one open space in each wheel. The Lyot stop is mounted on a reference surface in the input face of the filter module; the testing of the lyot stop has been discussed in Section 3.4. The slit wheel contains a selection of slits at different angles relative to the 0m slit, which defines the optic axis. A number of different angles are required to compensate for the field–rotation effect discussed in Section 2.3; a continuous choice of slit angle is obtained by selecting the slit closest to the required angle, then rotating the slit wheel about its centre to correct for the residual angle. The image of the slit will then be vertical on the array. The alignment of the 36.9° slit to be used with the echelle is most critical. It must lie radially in the wheel, since peak

transmission for a given wavelength is constant along a radius of the CVF. The slits are aligned using fiducial marks etched on the slits and slit wheel; this is done by eye using a 9× magnifying eyepiece. The angular accuracy achieved in this way can be found by measuring the angle of the slits using an angular graticule in the AT; the accuracy of this measurement was estimated to be 0.5°. A reference vertical was set–up using an arrangement of parallel bars and a straight edge and the slit angles measured relative to this. The results are shown in Table 3.3. These measurements were repeated during the calibration tests on the cold instrument (Section 3.11.1). Initially, the test of the echelle slit angle was attempted by aligning the AT crosswires to the edge of the CVF, assuming this to be radial, and checking the slit angle against them. An error in the echelle slit angle of 1.9° was found. However, the method of manufacturing the CVFs means that this is invalid as the edges are not in fact radial. The CVF is manufactured as a annulus and then sawn into quarters to provide the filters used, and it is the finite width of the saw that gives a non–radial angle to the filter edge. The widths of the slits are also given in Table 3.3: the 0.56mm slits are matched to the size of one pixel for the low–dispersion gratings, the equivalent slit width for the echelle is 0.45mm.

The f/convertor lens is mounted on the input face of the CVF/slit module; its alignment to this surface was tested by setting up the AT and laser and placing two targets in the module: one in the slit wheel and the other in the face of the module closest to the slits. With the module on the bench, and without the f/convertor lens, the targets are aligned to the AT by adjusting the position of the module. The angular alignment of the AT to the front reference surface where the lens will sit is checked and adjusted using the tilt controls on the AT. With the f/convertor lens in place, the displacement of the image of the target mounted in the slitwheel relative to the crosswire was seen to be less than 50μm .

A position around the filter wheel that can be used as a blank must be found to act as a safe position to prevent visible light from finding a path to the detector and to prevent any light from reaching the array when measuring the dark current and the bias levels. A position at 5050 steps was chosen manually, and the 1° (300 steps) offset either side also checked to allow for the possibility of stray light getting through. The three CVFs (1.3–2.5μm , 2.0–3.7μm and 3.7–5.5μm) were placed in the wheel with the

Slit Angle Results				
slit name	slit width (mm)	measured angle (°)	offset from vertical (°)	offset from design angle (°)
36.9m	0.45	117.0°	35.5°	-1.4°
-4m	0.56	157.0°	-3.5°	-0.5°
0m	0.56	154.0°	-1.5°	1.5°
4m	0.56	149.0°	3.5°	-0.5°
-4w	1.12	157.5°	-5.0°	1.0°
-7.4m	0.56	161.0°	-8.5°	1.1°
0n	0.28	153.5°	-1.0°	-1.0°

Table 3.3: The angle of the slits assembled in the wheel, relative to a defined vertical at 152.2°.

bandpass side closest to the slits so that the beam passing through is as near to focus as possible.

3.5.1 Assembly of the Camera/Collimator Module

The procedures outlined in this section apply equally to the short focal length and long focal length cameras. They described the internal alignment of the camera and collimator mirrors; the alignment to the optical casting is described in Section 3.6. The results quoted apply to the testing of the lflc, carried out at the Joint Astronomy Centre in February 1992.

The camera collimator is by far the most complicated of the modules, consisting as it does of five individual mirrors (Figure 3.1): a secondary collimating mirror in the first folding flat (FF #1), the primary collimating mirror (PCM) which reflects the beam onto

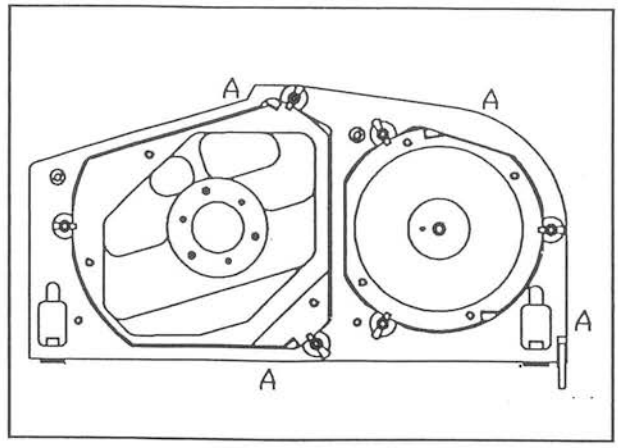


Figure 3.6: A rear–view of the cam/coll module showing the 'V' configuration for mounting the mirrors in the module. The mirrors shown are the two compound flats.

the first folding flat which in turn illuminates the grating, the second folding flat (FF #2) and the camera mirror which focusses light onto the detector. The mirrors are mounted on three locating pads by locating screws forming a 'V' configuration (Figure 3.6). A ZnSe field lens is located in the second folding flat. The alignment of these mirrors proceeds sequentially, beginning with the primary mirror and the first folding flat.

Before mounting the mirrors in the cam/coll module the following test set–up must be assembled: the module is placed on the casting and the AT aligned to the input port of the casting and the datums noted. The casting is clamped to the optical bench so that the module can be removed without disturbing the alignment. The mirrors are mounted in the module; it is crucial to the alignment that the retaining screws are tightened just until they touch the mirror and no further, as any stresses on the mirror will cause astigmatism and hence deterioration of the image quality.

With the mirrors in place the module was replaced on the casting and the location and alignment of the back face of the PCM measured. The angular tolerance on this measurement is 0.25mrad; the decentre tolerance is 0.025mm. The repeatability of this was checked by removing and replacing the module and checking the alignment. The final measurement of these quantities for the long focal length camera were 0.25mrad tilt

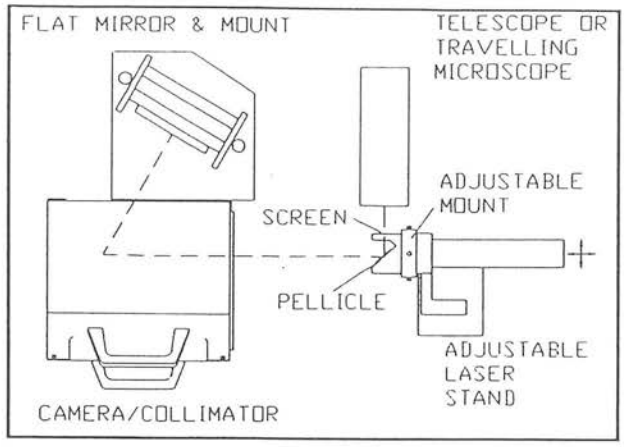


Figure 3.7: The measurement of the camera back focal lenght, which defines the slit position.

and 0.032mm decentre, a value which is exceeds the tolerance by 0.007mm. This excess was due to the height of the module above the casting being too great. However, this value for the decentre was found to be acceptable given that the estimated error on the decentre was 0.005mm and that the tilt was within tolerance.

The PCM is now correctly located. The camera back–focal length, the distance which defines the slit postion, can now be measured using the PCM and the first folding flat (FF #1) with a flat mirror mounted in the collimated beam in place of the grating. The test set–up is shown in Figure 3.7. The module is set up on the optical bench with the laser and beam expander aligned to the PCM and fully illuminating the secondary. The footprint on the dichroic mirror was checked with it angled at 14° to the vertical, so that the image of the pinhole is returned through the system and imaged on the screen of the laser pellicule holder. The distance of the laser from the pole of the PCM is adjusted, using the laser stage micrometers until the best focus for the return image of the slit was found and the slit distance from the PCM calculated. The theoretical value for the BFL is 50.0mm with a tolerance of 0.3mm; the measured value was 50.04mm$\pm$0.16mm

The system is now set up to fully illuminate the PCM and FF#1, so that alignment of the remaining mirrors can be carried out. With the dichroic in the collimated beam set vertical, the beam from the laser will illuminate all the mirrors and come to a focus at or near) the detector position. The rotation of the folding flats determines the footprint on the camera mirror, and the final position of the laser spot on the array. To check this alignment the camera mirror is replaced with a perspex blank which has the theoretical footprint of two concentric circles etched on it. There is a central obscuration, due to the hole in the primary, surrounded by the larger illuminated area. With the flats placed in their positions but without tightening the retaining screws, they can be rotated by moving a captive screw until the footprint on the perspex blank is correctly filled with light. Results for the short focal length camera showed both the obscured and illuminated areas to lie within the footprint to an estimated accuracy of $\pm$1mm. The long focal length camera profiles showed a consistent tendency for the dark obscuration to appear low, while the illuminated area was rather high. The origin of this problem was thought to be in the relative alignment of the compound flats, but there was not sufficient travel in the rotation of the flats to correct this effect.

The detector focal distance was measured by focusing the laser spot on a screen at the approximate focal distance from FF#2. The theoretical value for this position is 23.32mm and it was measured to be 23.06$\pm$0.24. This measurement will tend to be systematically low since it is measured with feeler gauges which will underestimate the distance if incorrectly used. The tolerance on this measurement is $\sim$0.1mm, since the focus of the ZnSe lens is strongly wavelength dependent and can be adjusted using the detector focus mechanism. As the error in the measurement is not so large as to imply a gross error in the optics, this value is acceptable.

The image quality of this pinhole image on and off axis was also checked at this point. As optical light was used, the spot sizes are not the final NIR spot sizes, however, the change in spot sizes of the on and off–axis beams is still a useful test, and the laser light can be used to check for astigmatism. For the short focal length camera the on–axis beam was measured to be $\sim 50\mu$m in diameter (estimating the 80% encircled energy) using a travelling microscope. The off–axis spots were measured by offsetting the microscope $\pm$8mm from the optic axis; the edge of a 58$\times$62 pixel detector with 76μm pixels will be

2.4mm from the optic axis. The image size at -8mm was as sharp as that on–axis and the size was again estimated to be 50μm ; at +8mm the spot size had increased to 100μm . Measuring the image quality with the long camera using a travelling microscope the spot sizes were found to be $\sim$66μm (two divisions on the microscope) for the on–axis beam and also for off axis beams at $\pm$4.5mm. During an early test, an astigmatic off–axis out–of–focus image was observed. The astigmatism was removed by loosening the retaining screw on the FF#2, which underlines the importance of the degree of force to be used when tightening these screws.

The cam/coll module can now be assembled on the casting, along with the filter and slit modules, the gratings and image rotator. The actual physical distance between the slits and the cam/coll must be measured to see if they agree with the empirical value for the required slit position of 50.04mm to within a tolerance of 0.3mm. The actual distance between these was 50.55mm, measured with feeler gauges between the slit module and cam/coll module and a depth micrometer to determine the slit position relative the face of the slit module. The position of the modules relative to each other are maintained by spacer bars; a new bar was machined to move the cam coll module 0.5mm closer to the slit module. The new slit distance was then measured to be 50.10mm.

3.6 Alignment of the slit, filter and camera modules to the casting

Each module sits on four mounting pads, machined to strict tolerances, which fully define the alignment of that module. The following test on the filter, slit and camera/collimator modules is to verify that the pads are correctly sized to align the modules to the casting within an angular tolerance of 0.25mrad and a linear tolerance of 0.05mm.

The angular alignment of the slit module is most critical, since light losses from the slits will reduce the instrument throughput. The decentre tolerance is influenced by the need to correctly align of the beam to the f–convertor lens; misalignment will also introduce vignetting. Any tilt of the filter module will cause scattering of light from the filters. This can act as a baffle, since the light which is not transmitted by the filter

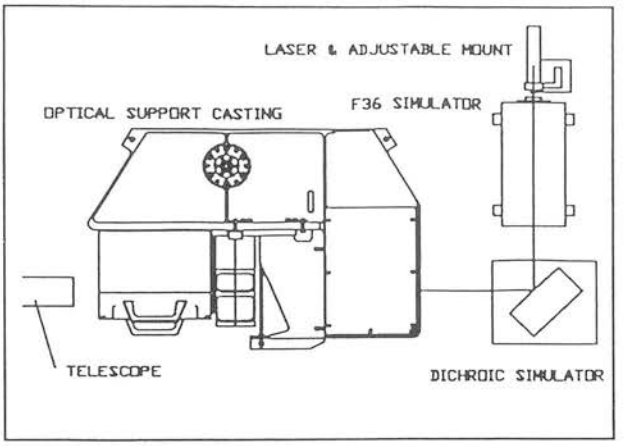


Figure 3.8: The test set–up for the aligment of the modules. The correctly illumination of the system is ensured by using the laser with the f/36 simulator and the dichroic mirror.

will be reflected back along the optical train and out of the cryostat, but ghosting will occur if the filters are misaligned. The tests were made with the modules mounted on the casting; they are fixed by three M6 screws which are firmly tightened. The laser and the AT were set up to be aligned to the casting which was clamped to the optical bench. Figure 3.8 shows the modules and casting located on the optical bench, with the system illuminated by the laser, using the f/36 simulator and the dichroic. To check the module positions a target and a flat mirror which sit in the lyot stop position (filter module) or the f/convertor lens position (slit module) were used. For the cam/coll module, a large perspex target was machined in the shape of the camera mirror, and a large flat mirror used. The baseline between a given module and the AT coverglass was also found. The barrel readings on the AT when it is aligned relative to the casting are noted as datums.

When the AT is focussed on the target, the position of the crosswires will have been moved off centre to some degree depending on the initial alignment accuracy. Whether the crosswires are displaced above or below the centre of the target determines whether the adjustment of the pads will require removing material from them (if the crosswires fall below the centre) or adding a shim (if the crosswires are above the centre). To measure the quantity of material to be removed or added in this process, the AT is

Summary of Pre–cool–Down Alignment Results						
module	datum		measurement		decentre	tilt
	Dial A	Dial B	Dial A	Dial B	(mm)	mrad
cam/coll	0.27	0.645	0.29	0.635	0.022	0.16
slit	0.29	0.67	"	"	0.032	0.25
filter	0.28	0.7	"	"	0.036	0.25

Table 3.4: Alignment Measurement pre-first cool down. The tolerance on tilt is 0.25mrad and 0.05mm on decentre.

aligned to the target in the module and the new barrel position noted. The tilt of the AT relative to the module is also noted.

The tests on CGS4 showed that the cam/coll module was aligned without further changes to the pads, however the slit and filter modules were too low and consistently out of spec on the 'B' dial of the AT. The degree of tilt was acceptable for all three modules. A shim of 0.1mm was added to the filter module and of 0.14mm to the slit module. The test was repeated with the new pads and the modules were found to be successfully aligned (Table 3.4). As in all of these experiments, the repeatability of the alignment was checked by verifying the measurements after removing and replacing the modules a number of times.

3.7 The Gratings

With the optical train complete, the footprint on the grating and the movement of the orders of the laser at the detector position, *i.e.* the tilt of the detector relative to the

dispersion direction, can be measured. The dispersion direction defined by the grating rulings must be aligned so that a spectrum of a point–like source lies along a single row of the detector to ensure the maximum signal/noise on that spectrum. The tolerance determined for this was that the alignment should be within 0.1pixels, *i.e.* at an angle of 1.5mrad. The tilt of the detector is set to be 7.38° due to the reflection of the image by the two compound flats, so the laser orders should move past the detector position at an angle of 7.38° to the horizontal as the grating is rotated. The laser orders were observed by setting the crosswires of the AT graticule to be horizontal and rotating the grating by driving the motor. The angle that the dispersion direction made with this horizontal lines was estimated relative to the crosswires, and accepted as being 7°. As part of the cold tests, the true detector rotation on the array was determined accurately.

During this test the zero–order position in motor steps was found; the blaze angle (θ_{blaze}) for the 75l/mm grating is $5°23\hat{'}$, therefore the brightest order (n_{max}) is 4th order for the wavelength of the laser light, 632.8nm (for the 150 l/mm $\theta_{blaze}=17°27\hat{'}$ and n_{max} = 6). This order is easily identified and from it the absolute value of the number of motor steps from datum required to bring the zero order position to the centre of the detector can be determined. This is used in calibrating the gratings (Section 3.9.1). The footprint on the gratings were also checked; the footprints are symmetrically placed on the gratings, but undersized. The low–dispersion grating footprints appear circular, whereas that for the echelle is elliptical due to the anamorphic magnification (Section 2.3).

3.8 Post cool–down tests of alignment

It has been maintained (Chapter 2) that, since there is no differential contraction of the mirrors and support structure, the optical alignment of the warm instrument will be preserved on cooling to LN_2 temperatures. This can be verified by measuring the image quality of the images in the NIR when the instrument is cold. A further test is to check the optical alignment when the instrument is again at room temperature.

The first test was to verify the module positions before the optics were dismantled; the method is the same as that used in Section 3.6. A resumé of the results is given

in Table 3.5. The total decentre of the cam/coll module is 0.058mm, greater than the 0.05mm tolerance set on this measurement. However, given that the value cannot be determined to better than 0.005mm, the error on this decentre value will be $\sqrt{2}\times0.005$ ($\sim$0.007mm), so that the result is not significantly different from 0.05mm. With the laser and f/36 simulator set up, the collimation of the output from the IROT was investigated. The diameter of the pupil was seen to be 9.2 $\pm$ 0.25mm at the IROT output and 9.6 $\pm$ 0.25mm at 0.8m as measured with vernier calipers. Again these measurements are in keeping with those carried out before the cool–down. Correctly determining the f/36 simulator focus, and placing the casting there, is crucial to the success of this test; an early measurement gave a pupil diameter of 6mm at a distance of 0.7m due to misalignment of the f/36 focus and the casting. The quality of the images in the detector plane was investigated. An on–axis image produced a 50μm spot size (80% encircled energy) on the screen placed at the nominal detector position. By placing the screen $\pm$8mm off-axis, images were observed with spot sizes $\sim$100μm , and that the spot at +8mm was sharper than that at -8mm. Comparing these results with those found previously (Section 3.5.1) it is seen that similar spot sizes were observed both on and off–axis, but that the quality of the -8mm spot has deteriorated from the previous measurement in which it was comparable to the on–axis spot. Some change in the optics of the camera module has clearly occurred, but the images are no less acceptable than before.

When the modules were removed from the casting the beam at the lyot stop position was remeasured using a screen with the pupil size scribed on it. For both on and off-axis beams the pupil size and image quality were as before (Section 3.4). The cam/coll module was examined separately using the perspex blank in place of the camera module to check the footprint. The illuminated and obscured areas lay within $\pm$1mm of the etched theoretical footprint.

The results of these test have demonstrated that no significant deterioration in the performance of the optics has occurred during the thermal cycle. The measurable changes all occur within the tolerances. Thus this aspect of the mechanical alignment is successful.

Littrow mount and a grating with blaze angle θ_{blaze} the grating equation can be expressed

$$n\lambda = 2d \sin\left(\theta_{grat}\right) \cos\left(\alpha\right) \tag{3.4}$$

where n is the order, d is the separation of the rulings which is the inverse of the ruling spacing, g, $\theta_{grat} = \theta_{blaze} - \epsilon$, ϵ is the grating rotation angle and α is the off axis angle, 14° with CGS4. The dispersion relation (wavelength range per pixel) can be shown to be

$$\delta\lambda = \frac{2df}{nw} \cos(\theta_{grat}) \cos(\alpha) \tag{3.5}$$

where f is the camera focal length (150mm or 300mm), w is the pixel width (76μm). The grating equation is parameterised in the form:

$$n\lambda = A \sin(B[(s - C) + D(s - C)^2]) \tag{3.6}$$

where s is the number of motor steps and A,B,C and D are the constants to be determined. A set of initial conditions for these parameters can be estimated by comparing the theoretical and parametric equations. A is found from the grating ruling spacing, 75 l/mm or 150l/mm, the order and the off-axis angle,

$$A = 2d \cos(\alpha) \tag{3.7}$$

B is a scale factor to convert between radians, the physically meaningful quantity and motor steps, the measured quantity. The stepper motors used have 200,000 steps per revolution, thus:

$$B = \frac{\pi radians}{100,000 steps} \tag{3.8}$$

C is the physical offset of the zero order position from the grating datum, and can be estimated by observing the number of motor steps required to bring a zero-order image onto the central (31st) pixel of the array. The constant D introduces a second order

There are 16 steps per pixel with the 150mm focal length camera: this is given from the ratio of the pixel size in μm and the focal length of the camera in mm scaled by the number of motor steps per degree.

$$pixel_{steps} = \frac{w}{f}\frac{180B}{\pi} \tag{3.10}$$

To complete the grating calibration, it is necessary to provide the least square fitting programme with an array of line wavelengths, line positions in pixels and steps; the grating equation (Equation 3.6) should be constrained over the entire range of grating angles covered for observation at 1–5μm by using lines throughout the range and/or by observing low wavelength lines in higher orders. A similar procedure to that employed for determining C is used, but an arc line is selected and the grating set to the wavelength of that line. The estimated parameters for A,B,C and D at this stage close enough in reality to place an arc line on the array, but not on the 31st pixel as desired. An observation of the spectrum is then made, and the line centre (in pixels) measured using the automatic line fitting facility in the data-reduction software. Since all observations are made with the 0m slit the line will not appear vertically on the array, so the line fitting must be carried out on the central row of the slit, determined earlier. The required data are grating order, arc line wavelength, grating steps, grating angle and line centre in pixels. This is repeated for a number of arc lines. When sufficient measurements have been made ($\sim$ 10 lines), the grating equation parameters can be determined. The true wavelength at the 31st pixel can be found from the dispersion relation (Equation 3.5) and the shift in pixels between the observed line and the central pixel. The parametric grating equation is the relationship between this true wavelength and the grating steps, found by fitting the data using a Chi-square least squares fitting function. The estimated values for A,B and D and the measured value for C are used as a starting guess. The tolerance on the fit is that the residual between the calculated and measured wavelengths should be less than $\sim$ 3 pixels with the low dispersion grating, but may be larger on the echelle ($\sim$ 10). A typical set of data for the low resolution grating calibration is given in Table 3.6; the value of the parameters in this case were:

A=24.85175 μm B=3.2181$\times 10^{-5}$ radians step^{-1}

C=2670 steps D=1.28906×10^{-6} steps^{-2}

The residual between the fit and the observational data is ± 12 steps (less than one pixel). A plot of the fit against the experimental data with the residuals is shown in Figure 3.9.

The procedure for the echelle grating is essentially the same, with the one difference that it is not possible to determine C directly, since there is no power at zero order with the echelle due to the large blaze angle (63.5°). The echelle C is fixed with reference to C determined for the low–dispersion grating. Theoretically there should be 100,000 steps (180°) difference between the value for C for the echelle and for the low dispersion grating, as they are mounted back to back. This may not be the true value, but adopting this value for C works well in practise and has some physical basis. The remainder of the line observations and fitting proceed as for the low–dispersion grating, but with the echelle a much larger range of grating angles can be used, which will define the curve thoroughly although only a narrow range of grating angles close to the blaze angle will be used in practice (~63.5°±4°). Typical values for the echelle data are

A=60.94444 μm B=3.14103×10^{-5} radians steps^{-1}

C=102729 steps D=3.20451×10^{-7} steps^{-2}

A plot of this fit against the data with the residuals is shown in Figure 3.10; the residual was ±4 steps.

3.9.2 Optimum Order for the Echelle

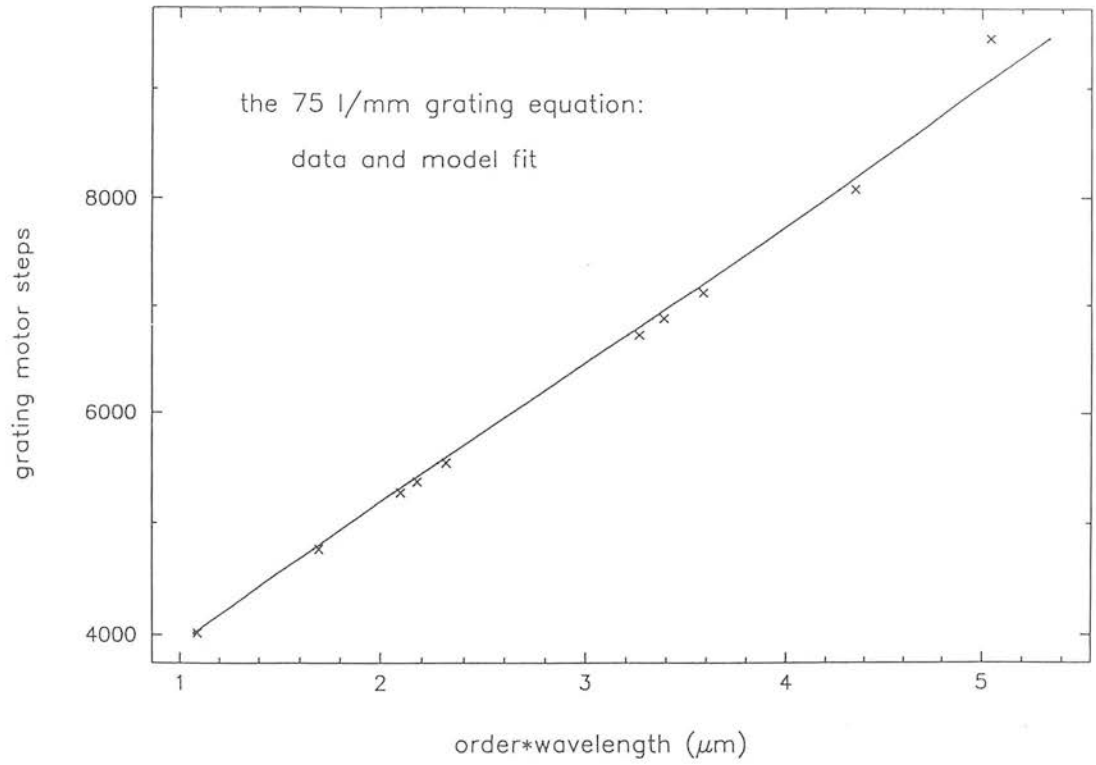
As previously stated, the maximum efficiency for the echelle is found when the grating angle is closest to the blaze angle (~ 63.5°) which is wavelength dependent. However it is not always possible to make these angles equal, since orders are discrete. It is necessary to determine the optimum order (n_{max}, the order at which the grating efficiency is

Wavelength μm	Order	Dispersion μm /pixel	motor steps	pixel	Wavelength at pixel 31
1.6945	1	0.00654	4768	27.48	1.6754
1.6945	2	0.00325	6875	27.96	1.6844
1.7920	2	0.00325	7120	34.06	1.7820
2.0992	1	0.00654	5271	33.69	2.0816
2.3140	1	0.00653	5537	34.00	2.2944
1.0887	1	0.00655	4018	33.50	1.0734
1.0887	2	0.00327	5368	34.94	1.0759
1.0887	3	0.00217	6723	32.75	1.0850
1.0887	4	0.00162	8085	34.00	1.0839
1.0087	5	0.00128	9458	35.00	1.0836

Table 3.6: Experimental data for the low resolution grating

Order	Dispersion $\times 10^{-4}$ μm /pixel	motor steps	pixel	Wavelength at pixel 31
10	15.0	111709	29.3	1.6970
20	6.487	121518	27.8	1.6966
25	4.505	127207	27.6	1.6960
30	2.910	134113	30.1	1.6948
31	2.600	135769	31.4	1.6944
32	2.282	137581	34.2	1.6938
33	1.949	139615	37.6	1.6932
34	1.584	141990	40.2	1.6922

Table 3.7: The data and fit for the echelle grating; the wavelength used was 1.6945μm in the orders listed.



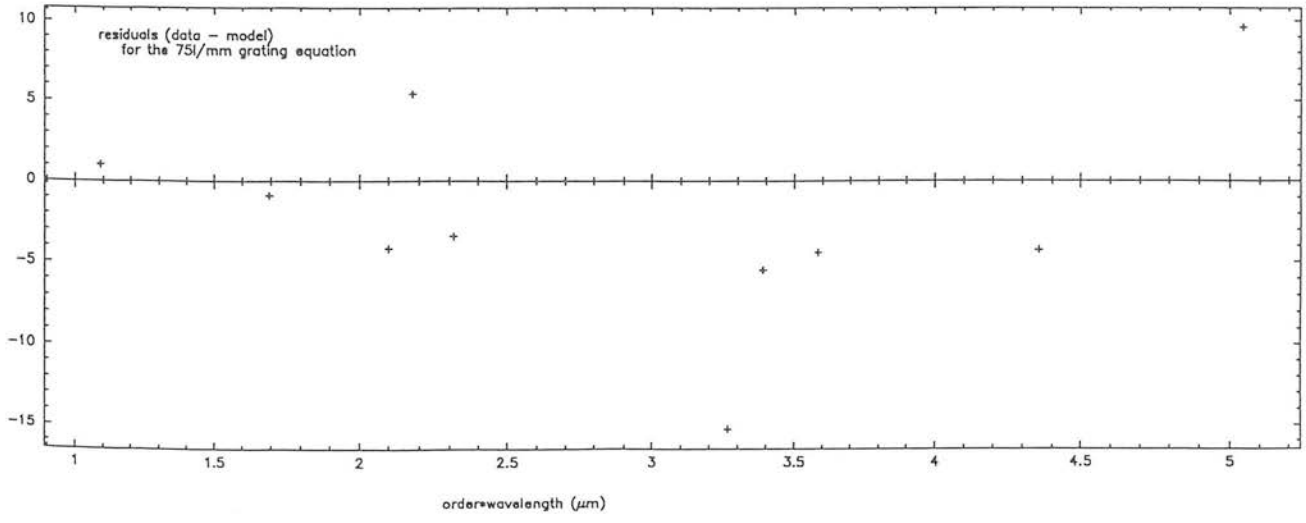
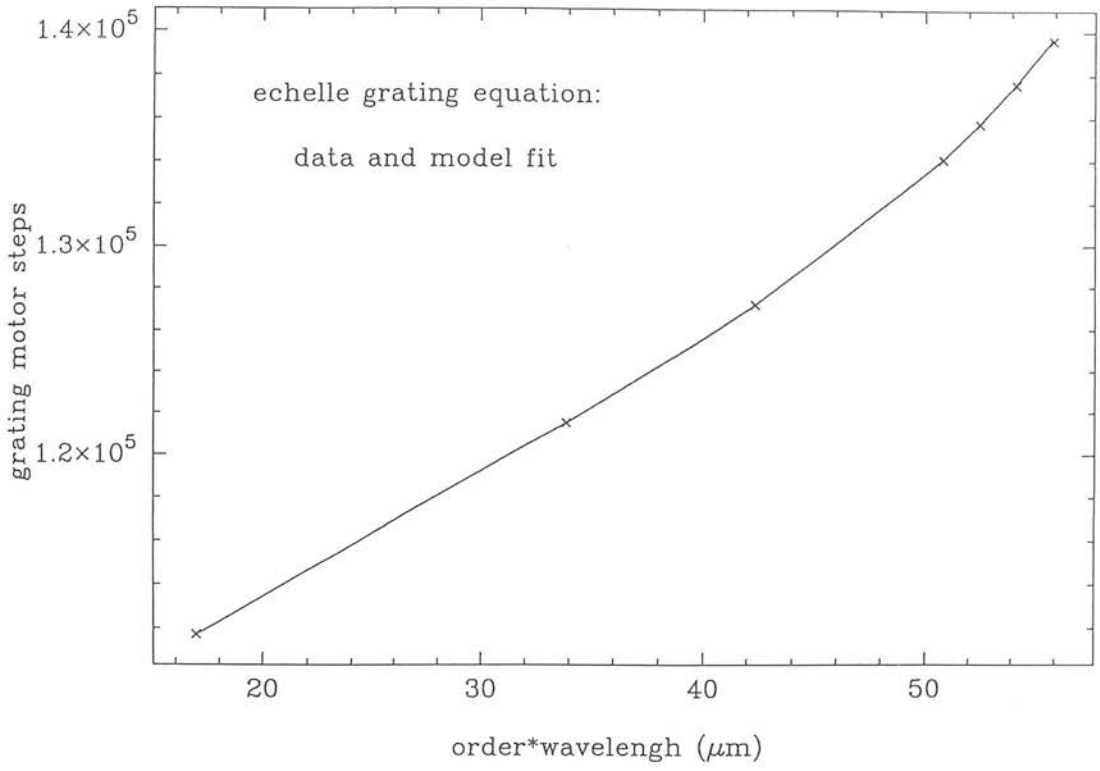


Figure 3.9: The data, model fit and residuals for the 75l/mm grating equation.



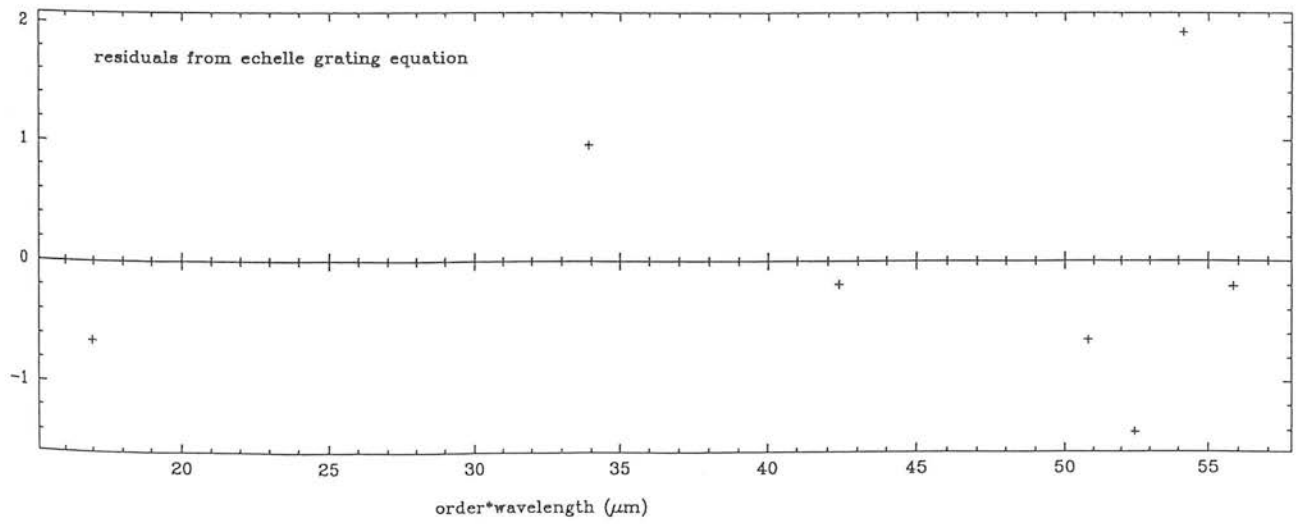


Figure 3.10: The data, model fit and residuals for the echelle grating equation.

greatest) for all wavelengths and ensure that this order is automatically chosen by the instrument control software whenever a wavelength is selected on the echelle. This can be done in two ways: either the order giving a grating angle closest to the blaze angle can be found by comparing the difference between θ_{grat} and θ_{blaze} for all orders, or by fitting a function which will give the optimum order for a given wavelength. The latter approach is purely empirical but the former is complicated by the fact that the blaze angle is a function of wavelength, changing by $\sim 1^{\circ}$ over the wavelength range, and by the fact that the degree of vignetting increases with the angle away from blaze. This latter effect is hard to quantify theoretically, and was the strongest reason for adopting the empirical approach. Determining the optimum order is very simple: a continuum source is selected and the grating wavelength set. The optimum order is selected by the software from the initial guess at the blaze angle. The level of signal received at the array can be found to sufficient accuracy for these purposes by noting the signal registered for each output channel of the array on the data acquisition screen during a Movie observation. By manually selecting the orders on either side of this estimated optimum order it can be seen if the signal increases or decreases. Since continuum sources are being used, any wavelength can be selected, to cover the entire range of grating angles. The relationship between optimum order and wavelength is then fitted with a polynomial of the form

$$n_{max} = d_0 + d_1\lambda + d_2\lambda^2 + ...d_n\lambda^n \tag{3.11}$$

The data obtained are given in Table 3.7, with the plot of fit against data shown in Figure 3.11. The smallest residuals were found for a 4th order fit with coefficients:

d_0=128.985	d_1=-111.68
d_2=44.658	d_3=-8.32
d_4=0.583	

Clearly the tolerance on this measurement should be that the residuals at all wavelengths are less than 0.5 so that the correct order is always picked, however, no fit to the data could be found to satisfy this requirement. This is a problem arising from

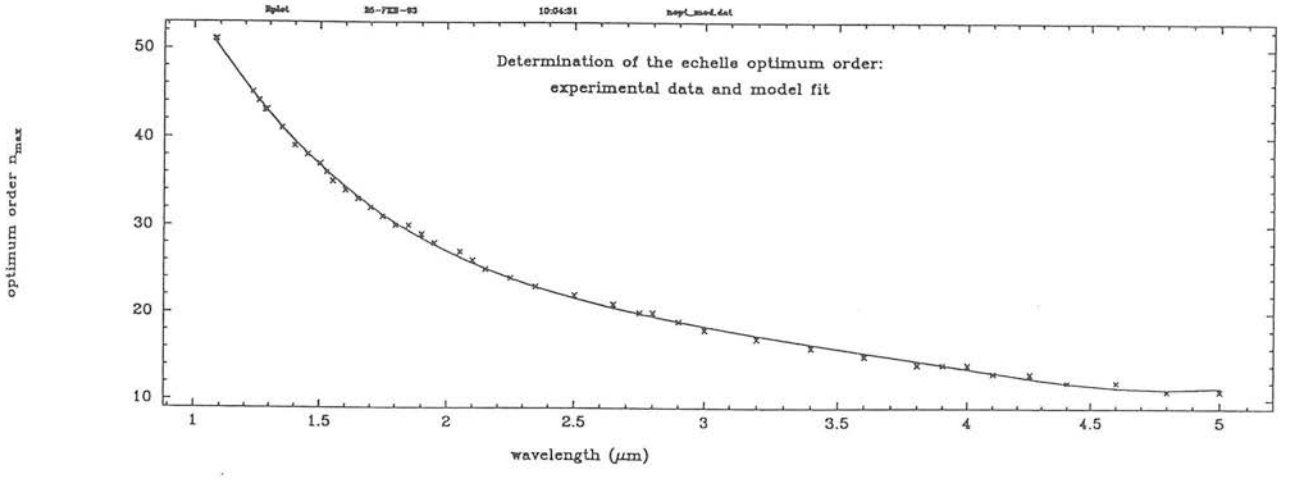


Figure 3.11: The experimental data for the determination of the optimum order for the echelle grating; the model fit is shown also.

Wavelength (μm)	Optimum Order
1.080	51
1.083	51
1.086	51
1.233	45
1.257	44
1.282	43
1.290	43
1.350	41
1.400	39
1.450	38
1.500	37
1.525	36
1.550	35
1.600	34
1.650	33
1.700	32
1.750	31
1.800	30
1.850	30
1.900	29
1.950	28
2.050	27

Wavelength (μm)	Optimum Order
2.100	26
2.150	25
2.250	24
2.350	23
2.500	22
2.650	21
2.750	20
2.800	20
2.900	19
3.000	18
3.200	17
3.400	16
3.600	15
3.800	14
3.900	14
4.000	14
4.100	13
4.250	13
4.400	12
4.600	12
4.800	11
5.000	11

Table 3.8: The experimental data used to determine the optimum order

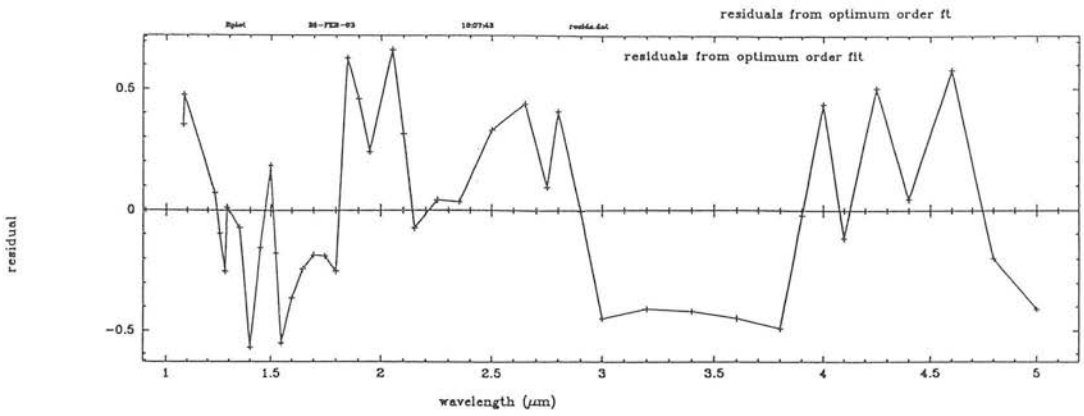


Figure 3.12: The residuals for the optimum order for the echelle grating. For those wavelengths where the value is greater than 0.5, the incorrect order will be chosen.

the discrete nature of the orders. Figure 3.12 shows that there are some wavelength ranges for which the residuals are greater than 0.5; they can be identified as the regions around 1.25μm , 2.0μm and 4.5μm . To illustrate the problem with an example, given a wavelength of 2.1μm and a blaze angle of 65.08°, orders 26 and 27 are equidistant in angle (2.33°) from the blaze angle. Uncertainty about the exact blaze angle at a given wavelength and the vignetting effects means that the optimum order cannot be chosen theoretically. It can, however, be tested empirically, and the results used to determine the parameters for Equation 3.11. Unfortunately, this will occur for a number of wavelengths, and we cannot test them all. It has been left to the observer to be aware of the wavelengths at which this problem arises, and to verify the optimum order quickly before relying on the chosen order. Of course, it may be decided that throughput can be sacrificed to higher resolution for observations of bright objects, in which case the optimum order serves only as a guideline.

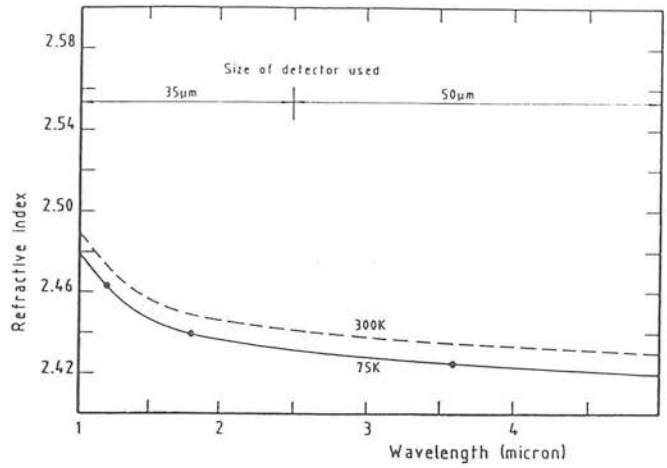


Figure 3.13: The change in refractive index of ZnSe with wavelength.

3.10 Focus Curves

The camera mirrors include a ZnSe field lens whose focal length is strongly dependent on wavelength; the refractive index of ZnSe is shown in Figure 3.13. Moreover, the high speed of the cameras (f/1.35, f/2.70) means that any small error in the focal length will dramatically affect the image quality: ray tracing shows that the spot size will increase from 20μm to 80μm for a 0.1mm error in the focus. The tolerance on the focal length is therefore set at 0.02mm. The theoretical values for the focus for a given wavelength were determined by ray–tracing with the CODEV package (Table 3.8); a general fit to these data is of the form:

$$focus = f_0 + f_1\lambda + f_2\lambda^{\frac{1}{2}} + f_3\lambda^{\frac{1}{3}} + f_4\lambda^{\frac{1}{4}} \tag{3.12}$$

the theoretical values for the coefficients are given in (Table 3.9).

Determining the focus curve for wavelength greater than $\sim$2.5μm is not feasible since the thermal background dominates so completely that no emission lines can be detected above it. Therefore, the theoretical focus curve is assumed and modified with a scale and offset which will be found from the experimental data.

$$focus = \mathrm{A} \times [f_0 + f_1\lambda + f_2\lambda^{\frac{1}{2}} + f_3\lambda^{\frac{1}{3}} + f_4\lambda^{\frac{1}{4}}] - \mathrm{B} \tag{3.13}$$

Wavelength μm	BFL (mm)	error (mm)
1.000	16.02	0.00
1.100	16.06	0.04
1.600	16.15	0.13
2.200	16.19	0.17
3.400	16.22	0.20
3.800	16.23	0.21
4.400	16.24	0.22
5.000	16.24	0.22

Table 3.9: CodeV data for focal length versus wavelength.

The focus is measured by taking arc line spectra at a range of wavelengths with a range of foci, all of which can be "automated" using the Exec system. The FWHM of the lines are measured using the automatic line fitting routines in the data–reduction; the widths will pass through a minimum value (theoretically one pixel) at the true focus. A sampling factor of 10 was used to ensured that the line profile is well defined; the centre can be determined to an accuracy of 0.1pixel. Since the focus changes fastest the 1–2.5μm region where arc lines can be measured, constraining the curve in this region and extrapolating gives a good approximation to the focus at longer wavelengths (Figure 3.13). Once A and B were determined, and the errors on the fitted curve were reduced to 0.02mm, it was verified that the same focus curve would provide in–focus images on the echelle. Typical examples of the values found for A and B were:

$$A = -0.12427 \qquad B = -1.8719$$

An important test of the optics is that the spatial and spectral focus points should be the same. If this were not so then this would be evidence for astigmatism. Using the

Parameter	fit to CODEV values
f_0	-71.7142
f_1	-4.0039
f_2	165.7249
f_3	-671.0576
f_4	581.0516

Table 3.10: CodeV values for the parameters controlling the focal length

focus given, the FWHM of both the profile along the slit and the profile along a column reached minima at the same focus to within the tolerance of 0.02mm set on this value.

3.11 Detector Rotation and Translation

The width of the slit has been chosen to exactly match the pixel size for the 58×62 pixel array to optimise the signal to noise ratio of the observations. To preserve this advantage, it is necessary to align the detector rows accurately with the dispersion direction (*i.e.* orthogonal to the grating rulings), else the light from a stellar source is spread over more than one row. The tolerance set on this is 1.5mrad (0.1 pixel across the length of the array). Moreover, with this optical configuration, the spectra are not fully sampled, as they would be if the image of the slit fell over two pixels, so that a method of sampling the spectra is required, to accurately remove sky emission and absorption features and to define the spectral line shape for unresolved lines. To this end, the detector is mounted on a motor–driven mechanism which can move the detector in the dispersion direction, and which also focuses the detector. The determination of the focus position has been discussed in Section 3.10.

The translation mechanism is able to move the detector up to two pixels widths in

the dispersion direction, and can be positioned with an accuracy of 0.1pixel to allow sub–pixel sampling (see Section 3.14). The "over–sampling factor" is the number of positions measured during an observation; in practise each position is a separate integration after which the array is read–out and the detector translated. For example, an observation made with an oversampling factor of 10 would consist of integrations at positions 0.0,0.1,0.2...0.9pixels. This is a larger factor than would commonly be used for astronomical observations, but which was used extensively during tests of the grating equation *e.t.c.*, where the line centre was to be accurately determined. For astronomical observations, 4 integrations at positions spread over 2 pixels provides a good compromise between adequately defining the line shapes, allowing the longest on–chip exposure times before the variation in OH lines requires a sky observation to be made (Chapter 4) and eliminating the effects of the bad–pixels, which is a by–product of sampling over two pixels.

The detector rotation, the angle between the rows of the detector and the dispersion direction, was measured when the instrument was first assembled (Section 3.7). For the tests of the cold instrument, the angle on the array made by the dispersed light from a pin–hole in the slit wheel, illuminated by a continuum source, can be found by fitting the line profile along the slit at either end of the array and finding the rows on which the peak flux falls. The result of this test was that a rotation of 0.8° was required. The detector rotation can only be adjusted by the use of micrometer screws mounted on the array housing. This means that it is not possible to adjust the rotation when the instrument is cooled; the correction must be made when the instrument is next warm. This is a rather troublesome problem, as determining the sense of rotation is far from trivial. The observation image made by the computer undergoes a number of "reflections" in the electronics during the read–out and in the reduction software before the measurement of the angle is finally made. An empirical method for determining the sense proved most successful as an aid to correcting for the numerous transformations that the data frame goes through. In zero–order, an image of an asymmetrically shaped mask can be made and placed at the input to the cryostat (near a focal plane). The effect of rotating this will be the same as the rotation of an astronomical object and the sense of rotation can thus be found. Of course, we still got it wrong!

Parameter	CODEV values
s_0	-0.138
s_1	0.515
s_2	-0.036

Table 3.11: Theoretical coefficients for the slit rotation equation from CODEV.

3.11.1 The Slit Rotation Equation

The projected angle of the slit on the array relative to the array axis, as defined by its rows and columns, is referred to as the slit rotation on the array and is caused by the variation in the out–of–plane angle with distance vertically on the slit (Section 2.3). The dependence of this rotation on wavelength has to be calibrated and corrected for by the automatic selection of a slit angled correctly to the optic axis for the grating angle chosen. Section 3.5 describes the choice of angled slits available in the slit wheel; angles between the discrete set in the wheel are obtained by rotating the slit wheel about its central axis. Once again, CODE V was used to fix the theoretical form of the slit–rotation curve

$$\tan(\theta_{slit}) = s_0 + s_1 \tan(\theta_{grat}) + s_2 \tan^2(\theta_{grat}) \tag{3.14}$$

the values for s_i are listed in Table 3.10.

The tolerance on determination of the slit equation was that the slit should be aligned to within 0.1pixels, an angular tolerance of 3mrad. Experimental values for these constants are found by determining the slit for a 1–5μm range of wavelengths. Again, multiple orders may be used since the dependence is on grating angle rather than wavelength. Initially this test was carried out by using the 0m slit and measuring the tilt on the array by fitting arc lines at either end of the slit and finding the angle between them. However, it was noticed that with the echelle fixing the slit rotation in this way

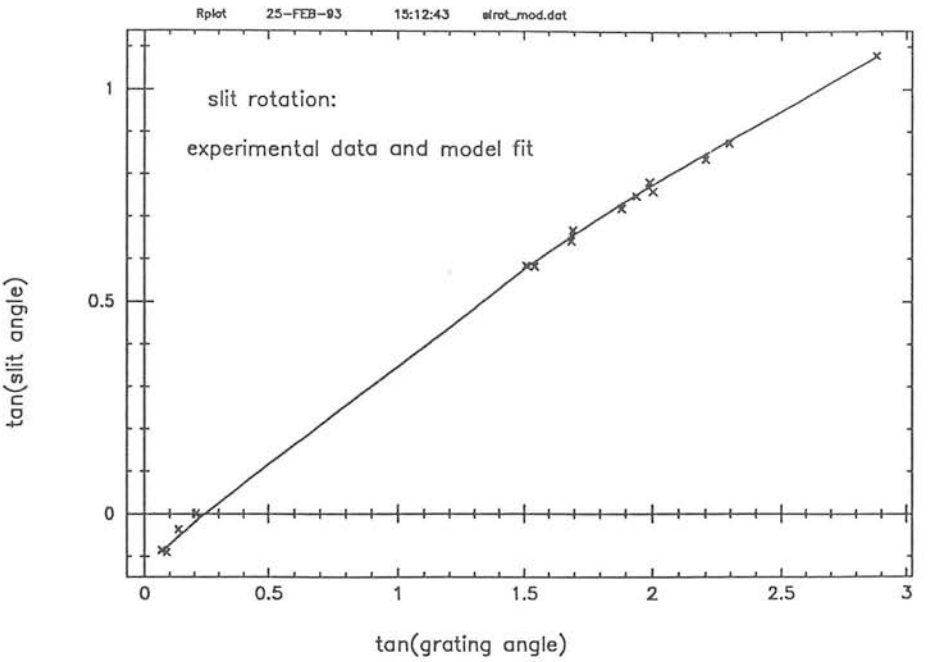


Figure 3.14: The fit to the experimental data for the determination of the slit rotation equation. The residual from this slit produces a shift of 0.007pixels from one end of the slit to the other.

was inadequate. The offset between the projected angle on the grating and the physical angle relative to the 0m slit is not simply the arithmetic difference between these angles, so that it cannot be corrected by rotating the slit to the measured angle. This effect is so small as to be negligent on the low–dispersion grating, but large enough on the echelle to cause noticeable discrepancies. The new method that was adopted to find the slit angle is to chose an arc line, select either grating, and, using Movie, adjust the slit–steps "manually" until the image appears vertical on the array (which can be judged quite accurately ($\sim$ 10 steps) by observing the illumination in the surrounding pixels at both ends of the slit) and note the angle. This is the true slit angle required at the grating angle selected. One advantage that we have in determining the slit angle is that the curve is the same for both the echelle and low–dispersion gratings, so the curve is constrained for θ_{grat} from 0° to 70°. The data are listed in Table 3.11, with a plot of the data and fit given in Figure 3.14.

Wavelength μm	Order	θ_{grat} (°)	nominal slit angle °	experimental slit angle °
1.6945	30	56.457	36.9	30.3
-	31	59.458	-	33.7
-	32	62.754	-	36.8
-	33	66.469	-	41.2
-	34	70.850	-	47.2
2.0992	25	59.366	-	32.7
	26	63.492	-	37.2
1.0888	47	57.033	-	30.2
1.3626	40	63.344	-	38.0
3.3676	16	62.059	-	35.7
2.3140	24	65.582	-	39.9
1.0888	2	5.009	-4.0	5.2
1.6945	1	3.899	-4.0	-4.9
-	2	7.795	0.0	-2.1
-	3	11.708	0.0	0.0

Table 3.12: The experimental data for the slit rotation equation. Lines were measured on both the echelle and the 75l/mm grating.

Experimental values for the slit equation coefficients were determined to be:

$$s_0 = -0.12427 \qquad s_1 = 0.52044 \qquad s_2 = -0.03579$$

The largest residual error with this equation is a shift of 0.007 pixels over 30 rows, a negligible angular error of 0.2mrad. The value for s_0 is within about 10% of the theoretical value, whilst the other two agree to better than 1%, so that the form of the theoretical curve is excellent, but an offset is required for the experimental data. This could possibly be explained by an offset of the 0m slit relative to which all the other angles are defined. As well as calibrating the slit equation using the 0m slit, the true angular offset of the slits from the 0m slit had to be calibrated and this offset applied when the slit is set to the required angle. This is a repeat of the experiment carried out on the warm instrument, the results for which were listed in Table 3.3. However, the slits had been removed and replaced between these measurements so that the results should not be compared. The advantage of carrying out this measurement using lamp lines and the detector is the greatly increased accuracy. The uncertainty introduced by the line fitting error is typically 0.1° over the slit length, compared to an estimated error of 0.5° previously. An argon line at 1.6945μm was observed using each slit in turn; an oversampling factor of 10 was used to accurately define the line shape. The offsets found are listed in Table 3.12; the greatest error was -0.69 measured for the 36.9m slit used with the echelle.

3.12 Calibrating the CVFs

Order sorting the echelle is achieved by using a set of narrow band filters below 1.3μm and three continuously or circular variable filters (CVFs), with a spectral resolution of about 2%, in the ranges 1.3–2.3μm , 2.3–3.7μm and 3.5–5.0μm . The CVF passband is typically 0.04μm (for a line at 2μm), compared with the array coverage with the echelle of 0.016μm , so that if the CVF is not correctly calibrated, the line will be observed with greatly reduced transmission or not at all. Selecting an arc line on the echelle and then

Slit name	Offset
36.9m	-0.6889
-4m	-0.1340
0m	-
4m	-0.4339
-7.4	0.0439
0n	0.0347

Table 3.13: Slit offsets relative to the 0m slit determined during cold tests.

stepping the CVF until the number of counts in the line passes through a maximum, describes a parabola whose maximum represents the peak CVF transmission for the line, *i.e.* the CVF wavelength corresponding to the motor steps. A simple linear relationship describes the peak wavelength as a function of motor steps; the linearity of the CVF is quoted as 1% by the manufacturers and was verified experimentally. Hereafter, only three arc lines are used to define the CVF equations which take the form:

$$CVF_{steps} = scale \times \lambda + zero \tag{3.15}$$

Calibrating CVF3, whose wavelength range is 3.7μm –5.5μm , with the emission lines sources proved difficult, due to the high thermal background and lack of bright arc lines. The final calibration of this segment was carried out using atmospheric features at 4.8135μm and 3.7086μm after the instrument was installed at UKIRT and checked with a bright Ar line at 3.9909μm ; the values obtained differed significantly from those found using arc lines, reflecting the difficulty of this calibration. A summary of the results obtained is given in Table 3.13.

Segment	Scale	Zero
CVF1 1.3–2.5μm	20947.5	-16164.6
CVF2 2.3–3.7μm	-13519.0	88939.4
CVF3 3.7–5.5μm	13320.1	20754.5

Table 3.14: CVF equations.

3.13 Image Quality

A true measure of the image quality of the instrument can only be obtained with the cold instrument using infrared emission line sources. The theoretical spot sizes were defined as 30μm when the design of the instrument and the mechanical tolerances were specified. For the 76μm pixels, the FWHM of the slit images should be no greater than 1.08 pixels.

The image quality at the centre of the array, and the degradation of the images at the extremes of the array were measured using both the low dispersion gratings and the echelle. Figure 3.15 shows an example of the arc spectra obtained with the low–dispersion grating using an oversampling factor of ten. The FWHM the strongest line at the centre of the array, the 1.6945μm argon line, is 1.02 pixels for row 30, and 1.03 pixels and 1.01 pixels at the extreme ends of the slit, rows 17 and 43 respectively. Figure 3.16 shows arc spectra taken at three positions on the echelle, the central arc line was a 2.0992μm line from the argon lamp, those at the extreme edges of the array are both from the 1.6945μm argon line. The change in FWHM for these three lines is 0.03pixels. The good image quality from the lab test is also demonstrated by the observations made at the telescope, presented in Chapters 5 and 6. The unresolved OH emission lines used to calibrate echelle spectra of the Galactic centre have an average FWHM of 37.9$\pm$1.0km s^{-1} , compared with the theoretical resolution of 37.2km s^{-1} . The line width of the HI Brγ line from Hubble 12, which was also unresolved, was

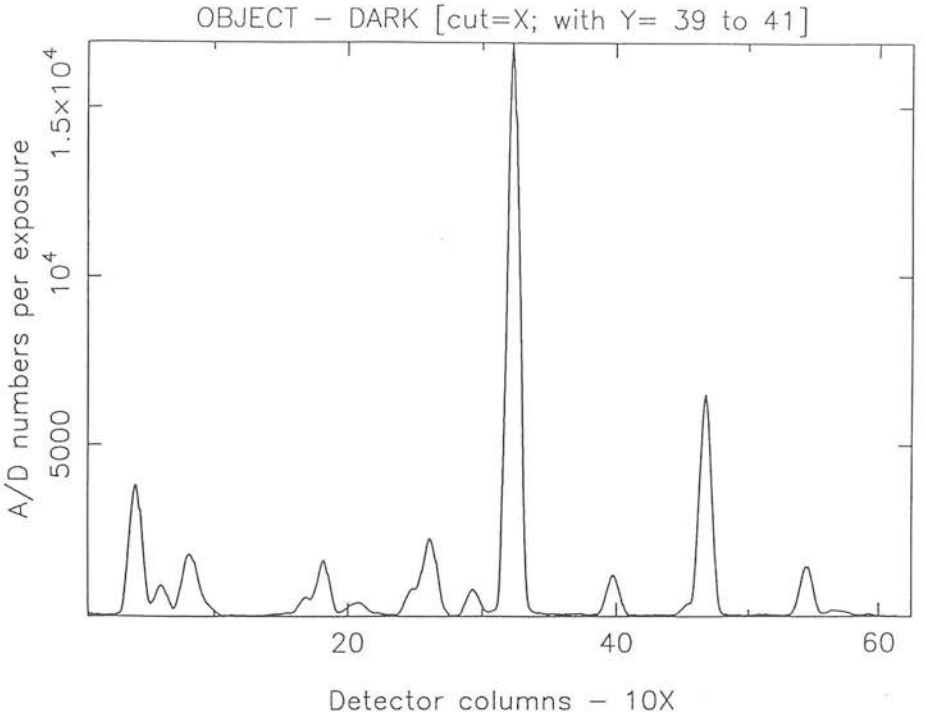


Figure 3.15: A low–dispersion spectrum of the argon lamp; the FWHM of the 1.6945μm line at pixel 32 is 1.02pixels at the central row. The variation in width of the line at the extremes of the slit are $\pm$0.01pixels.

0.0067μm compared with the theoretical value of 0.0064μm .

3.14 Reproducibility of calibration and positioning

A number of tests were carried out to ensure that the motor positioning of the gratings, translation and focus drives was sufficiently accurate over long periods to ensure the integrity of the calibration between "warm–ups". These test were carried out overnight using the EXEC system to perform large numbers (100s) of observations.

The grating tests were performed by repeated observing the same arc spectrum, datuming the motor between observations. The line was observed using an oversampling factor of ten, to allow the line centre to be determined to within 0.1pixels. Line fitting

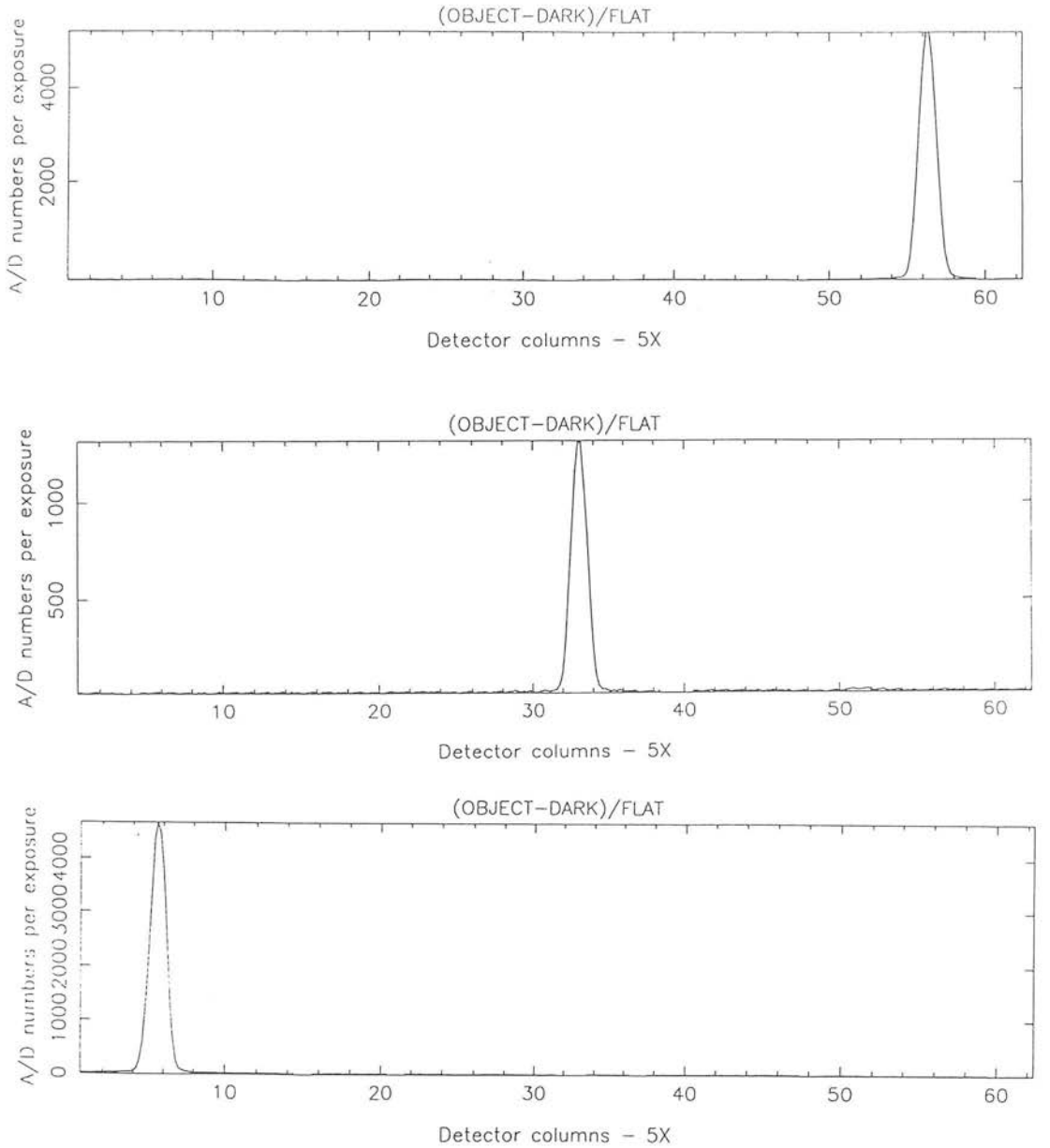


Figure 3.16: Three spectra of lines on the echelle show that the FWHM of the lines changes by only 3% across the array. The line at centre is the 2.0992μm argon lamp; those at the extremes are the 1.6945μm line.

of the resulting arc spectrum was carried out in real–time using the data reduction software. To make the test as realistic as possible, it was estimated the grating would be re–datumed approximately 200 times per semester, so this was the number of repeat observations. The tolerance for the grating reproducibility was that the line centre should remain within half a pixel of its starting position (*i.e.* the motor has only lost between 5–10 steps). The detector translations repeatability was toleranced to be accurate to within 0.05pixels, since an oversampling factor of ten may be in use. During tests the translation assembly was found to be repeatable to within 0.02 pixels. An effective drift was seen during the course of these measurements, thought to be due either to a thermal effect resulting in some expansion of the worm wheel, or due to wear on the wheel. Due to the pressure of time, these effects were never eliminated, and their origin was not determined. Figure 3.17 shows this effect - the scatter due to an error in positioning is superposed on a DC drift. These are the results from a test which was carried out on the combined grating and translation repeatability. A five minute wait was included between each observation to try and eliminate thermal effects. Two arc spectra were observed alternately with an oversampling factor of ten and the resulting line centre in pixels vs. observation number is plotted. The rms scatter from the two sets of data is 0.2pixels.

3.15 Conclusion

This chapter has covered the tests carried out on CGS4 from the initial alignment of the instrument to its final stages as a fully calibrated, common user instrument. The adopted mechanical alignment method has been described, and data given which demonstrate that this approach can be used when tight tolerances are required. Moreover, it has been shown that, as predicted, the alignment is preserved during the cooling and warming of the instrument and that the modules can be freely exchanged as the alignment is reproducible.

The calibration of the instrument has been achieved, and the tests on repeatability of the positioning of the grating and translation mechanism have proven that the calibration will be secure for the likely length of observation lifetime of the instrument. The stated

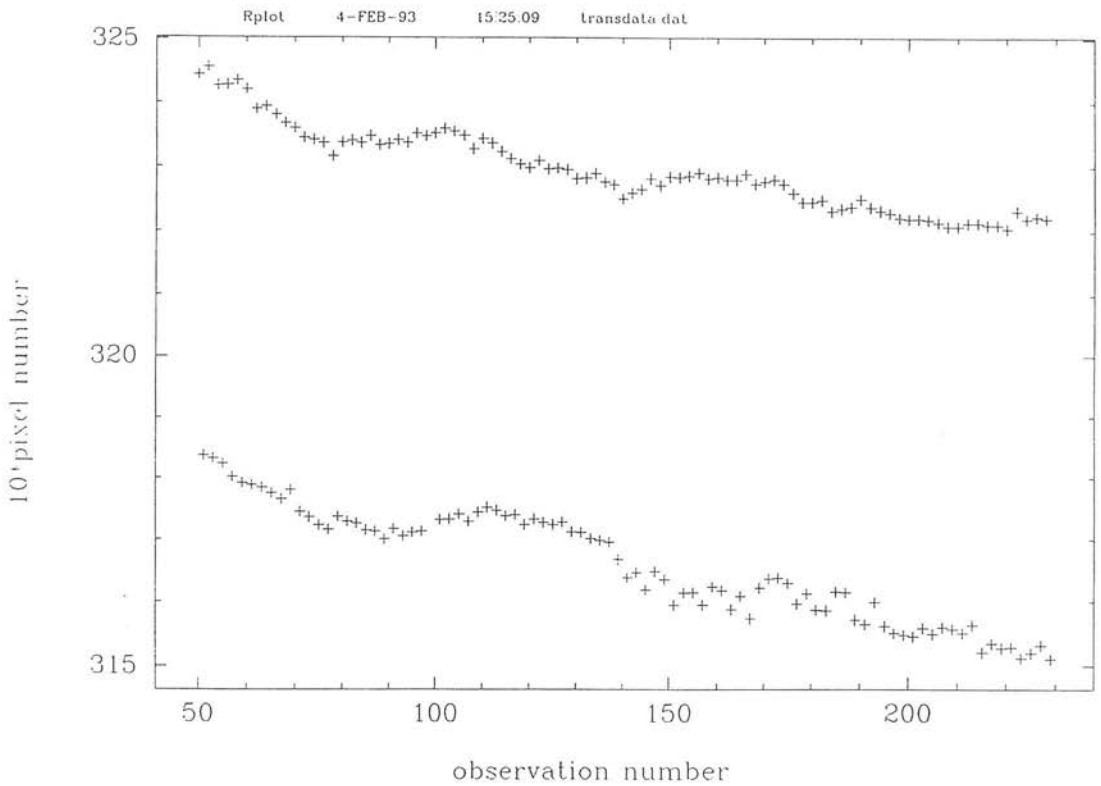


Figure 3.17: Repeatibility of the translation mechanism and grating drives; the rms scatterer on the positioning was 0.2 pixels.

aim, of providing a "black–box" into which the observer feeds the wavelength and position angle of the source, has been achieved but with one exception: the need to be aware of the possibility of obtaining an order other than the optimum order when using the echelle.

The results of the tests of image quality for the spectrometer in the NIR give the proof that the design, manufacture and alignment of the optical components of the instrument has been successful. The measured FWHM for a spectra of both calibration and astronomical sources, for the echelle and the low–resolution gratings, and at all points on the array, are reliably one pixel, within a 10% tolerance.

Chapter 4

Sky Noise from Atmospheric OH Emission

4.1 Introduction

Emission from the atmosphere is an abiding problem for the infrared astronomer. At wavelengths longward of $\sim 2.3\mu$m , the thermal contribution from the telescope and surroundings dominates and observations become quickly background limited. The emission in the region 1–2.5μm is principally non–thermal radiation from excited molecules in the upper atmosphere (85–100km), the so–called infrared nightglow, the most important component of which is the non–thermal emission from vibrational–rotational transitions of the hydroxyl radical, OH. This spectrally structured emission also shows periodic variations in its intensity, both temporally and spatially. The OH Meinel bands were first identified by Meinel in 1950 and observed in the airglow for vibrational quantum number $v \leq 9$ (Meinel 1950*a,b*). Variations in the emission intensity are seen over both long and short timescales: long term fluctuations are due to diffusion of the minor atmospheric constituents which alters the number density of reactants for the key equations that excite the OH (Equations 4.2 & 4.3); shorter period variations, of the order of a few minutes to an hour, are connected with the passage of density and temperature perturbations through the upper atmosphere. These gravity waves can be excited by any phe-

nomenon affecting the atmosphere over short timescales (Hines 1974) *e.g.* meteorological effects such as thunderstorms (Taylor& Hapgood 1988) or man–made effects. Their passage causes changes in the reaction rates and column densities in the constituents of the upper atmosphere and a corresponding variation in the emission intensity. Both these phenomena are discussed more fully in Section 4.4. These systematic variations in the atmospheric background can impose a limit on the sensitivities achievable in the NIR. In the spectroscopic, low–background, read–noise limited regime the major gain in signal/noise is obtained by long integrations of up to tens of minutes. Over these timescales, variation in non–thermal background emission causes an apparent increase in the "noise" level unless sufficiently frequent sky monitoring is employed. This can be a fundamental limit to the sensitivity achievable with staring arrays. In this chapter, the nature and source of this emission and its variation is described and results on the effect on instrumental sensitivity are presented. Possible strategies for minimising this noise source are discussed. This work has been published in *Monthly Notices of the Royal Astronomical Society*, and was carried out in collaboration with Tom Geballe and Matt Mountain (Ramsay, Mountain & Geballe 1992).

4.2 Observations of the OH Emission Spectrum

The spectra presented here were observed at UKIRT in November 1989 using CGS4's predecessor, CGS2, which incorporates a linear array of 7 pixels. A complete set of filters ranging from $1 - 2.5\mu$m was used, with a resolving power of $\sim$250 at the short wavelength end of the J window to $\sim$800 at the long wavelength end of the K window. Higher resolution observational spectra can be found in Oliva & Origlia (1991); it should be noted that they quote wavelengths in air. Measurements of the sky were made in "stare" mode and two standard stars (BS718 & BS1543) were observed in "chop" mode to obtain flux calibration. The observations were dark subtracted.

4.2.1 Data Reduction and Results

Algorithms were developed which describe the signal reaching the detector for both the sky and stars, including effects due to the atmosphere and the instrument. The sky–subtracted star signal is the product of the instrumental profile with the atmospheric transmission and the stellar flux distribution, the latter of which we approximate to a blackbody. To obtain the instrumental profile alone a stellar spectrum was divided by a black–body calculated at the effective temperature of the star and normalised at 2.2μm . IRTRANS, the atmospheric model of Traub & Stier (1976), was used to select points at which the atmosphere is transparent (Figure 1.1). The instrumental profile was calculated at these points and the full instrumental profile for the window found by interpolating between them, assuming that it varies slowly with wavelength. Such a profile was produced for the J, H and K windows.

The sky signal is the product of the sky emission (thermal as well as non–thermal) and the instrumental transmission profile as determined above. Thus, by dividing the raw sky frames by the instrumental profile we obtain the spectra of the sky emission. For wavelengths at which the atmospheric transmission is ≤ 0.75 we set the flux to zero. The wavelength scale was fixed using an argon arc lamp, and identification of the lines achieved by comparison with the wavelengths quoted in Chamberlain (1961). The agreement between our measured wavelengths and those in Chamberlain is better than 0.002μm , after the latter were corrected to in vacuo wavelengths. The wavelengths identified in our spectra, and the transitions involved, are listed in Tables 5.1,5.2 and 5.3 for the OH lines. Note that the strong line at 1.27μm is not an OH line, but the $O_2(^1\Delta_g)$ line produced through chemical reactions at the same altitude as the OH, 80–100km. (For a discussion of the chemical production mechanisms see Howell *et al.* 1990).

The results are presented in Figures 4.1,4.2 and 4.3. The wavelengths are in vacuo. The Q,P and R branches of various vibrational transitions are clearly visible, particularly in the K window. The Q–branch is unresolved; each P and R transition is a doublet, due to angular momentum splitting, and these are sometimes resolved depending on the resolution of the filters and gratings used. See for example the K window spectrum

Measured Wavelength /μm	Transition	v"-v'	Intensity /photons s^{-1} arcsec^{-2}
1.0063	Q	9 – 5	3399.6
1.0263	Q	4 – 1	2276.3
1.0363	$P(1)bl$†	4 – 1	662.0
1.0393	$P(2)bl$	4 – 1	766.2
1.0422	$P(3)bl$	4 – 1	498.3
1.0466	$P(4)bl$	4 – 1	431.0
1.0503	$P(6)bl$	4 – 1	357.0
1.0703	*Rbranch*	5 – 2	572.8
1.0824	Q	5 – 2	676.9
1.0913	$P(1)bl$	5 – 2	572.8
1.0945	$P(2)bl$	5 – 2	647.2
1.1002	$P(3)bl$	5 – 2	431.4
1.1043	$P(4)bl$	5 – 2	357.0
1.1312	*Rbranch*	6 – 3	468.8
1.1430	Q	6 – 3	825.7
1.1524	$P(1)bl$	6 – 3	572.8
1.1573	$P(2)bl$	6 – 3	550.5
1.1636	$P(3)bl$	6 – 3	453.7
1.1696	$P(4)bl$	6 – 3	364.5
1.1980	*Rbranch*	7 – 4	528.1
1.2092	Q	7 – 4	758.7
1.2186	$P(1)$	7 – 4	476.0
1.2254	$P(2)$	7 – 4	550.4
1.2311	$P(3)$	7 – 4	379.3
1.2379	$P(4)$	7 – 4	275.2
1.2660	*Rbranch*	8 – 5	801.1
1.2897	Q	8 – 5	1019.1
1.3010	$P(1)bl$	8 – 5	591.0
1.3072	$P(2)bl$	8 – 5	602.5
1.3128	$P(3)bl$	8 – 5	305.6
1.3210	$P(4)bl$	8 – 5	*
1.3286	$P(6)bl$	8 – 5	*
1.3996	$P(2)bl$	9 – 6	*
1.4139	*Rbranch*	2 – 0	*
1.4326	Q	2 – 0	*
1.4501	$P(1)bl$	2 – 0	937.4
1.4582	$P(2)bl$	2 – 0	929.9
1.4662	$P(3)bl$	2 – 0	654.7
1.4818	$P(4)bl$	2 – 0	825.8
1.4873	$R(1) - R(7)bl$	3 – 1	1123.3

Table 4.1: The intensities measured for OH emission lines in the J band. * denotes a line present in the spectrum, but where the atmospheric transmission is poor. † *bl* denotes a blend of two or more lines.

Measured Wavelength /μm	Transition	v"-v'	Intensity /photons s^{-1} $arcsec^{-2}$
1.5056	Q	$3-1$	*
1.5170	$P_2(1)$	$3-1$	1082.9
1.5235	$P_1(1)$	$3-1$	3884.0
1.5319	$P(2)bl$	$3-1$	3556.2
1.5403	$P(3)bl$	$3-1$	2572.8
1.5524	$P(4)bl$	$3-1$	2259.9
1.5578	$R(2)bl$	$4-2$	1708.6
1.5622	$R(1)bl$	$4-2$	2483.4
1.5814	Q	$4-2$	6715.0
1.5946	$P_2(1)$	$4-2$	1038.2
1.6018	$P_1(1)$	$4-2$	3749.9
1.6059	$P_2(2)$	$4-2$	1634.2
1.6108	$P_1(2)$	$4-2$	4375.7
1.6217	$P(3)bl$	$4-2$	2960.2
1.6330	$P(4)bl$	$4-2$	1678.8
1.6434	$P(5)bl$	$4-2$	1574.5
1.6489	$Rbranch$	$5-3$	2021.5
1.6692	Q	$5-3$	6461.6
1.6839	$P_2(1)$	$5-3$	814.7
1.6891	$P_1(1)$	$5-3$	3466.8
1.6944	$P_2(2)$	$5-3$	1366.0
1.6992	$P_1(2)$	$5-3$	3839.3
1.7061	$P_2(3)$	$5-3$	1217.0
1.7101	$P_1(3)$	$5-3$	2706.9
1.7235	$P(4), R(5,4,3)bl$	$5-3, 6-4$	1649.0
1.7367	$P(5), R(2)$	$5-3, 6-4$	1917.2
1.7418	$R(1)bl$	$6-4$	1917.2
1.7487	$R(0)bl$	$6-4$	576.3
1.7639	Q	$6-4$	5821.0
1.7860	$P(1)bl$	$6-4$	*

Table 4.2: The intensities measured for OH emission lines in the H band.

Measured Wavelength /μm	Transition	v"-v'	Intensity /photons s^{-1} $arcsec^{-2}$
1.9635	$R(2)bl$	8 – 6	793.4
1.9685	$R(1)bl$	8 – 6	1031.6
1.9756	$R_1(0)$	8 – 6	1638.8
1.9845	$R_2(0)$	8 – 6	620.8
2.0003	Q	8 – 6	*
2.0218	$P_2(1)$	8 – 6	*
2.0270	$P_1(1)$	8 – 6	2162.8
2.0315	$P_2(2)$	8 – 6	1019.6
2.0400	$P_1(2)$	8 – 6	2710.5
2.0470	$P_2(3)$	8 – 6	*
2.0556	$P_1(3)$	8 – 6	*
2.0722	$P_2(4)$	8 – 6	*
2.0854	$P_2(5)$	8 – 6	573.1
2.0897	$P_1(5)$	8 – 6	811.3
2.1110	$R(2)bl$	9 – 7	853.5
2.1183	$R(1)bl$	9 – 7	883.2
2.1250	$R_1(0)bl$	9 – 7	1044.0
2.1333	$R_2(0)bl$	9 – 7	496.3
2.1518	Q	9 – 7	2201.2
2.1704	$P_2(1)$	9 – 7	409.1
2.1792	$P_1(1)$	9 – 7	1533.7
2.1870	$P_2(2)$	9 – 7	569.2
2.1938	$P_1(2)$	9 – 7	1550.6
2.2042	$P_2(3)$	9 – 7	639.7
2.2108	$P_1(3)$	9 – 7	1461.3
2.2225	$P_2(4)$	9 – 7	492.3
2.2309	$P_1(4)$	9 – 7	1081.7

Table 4.3: The lines intensities measured for OH emission lines in the K band.

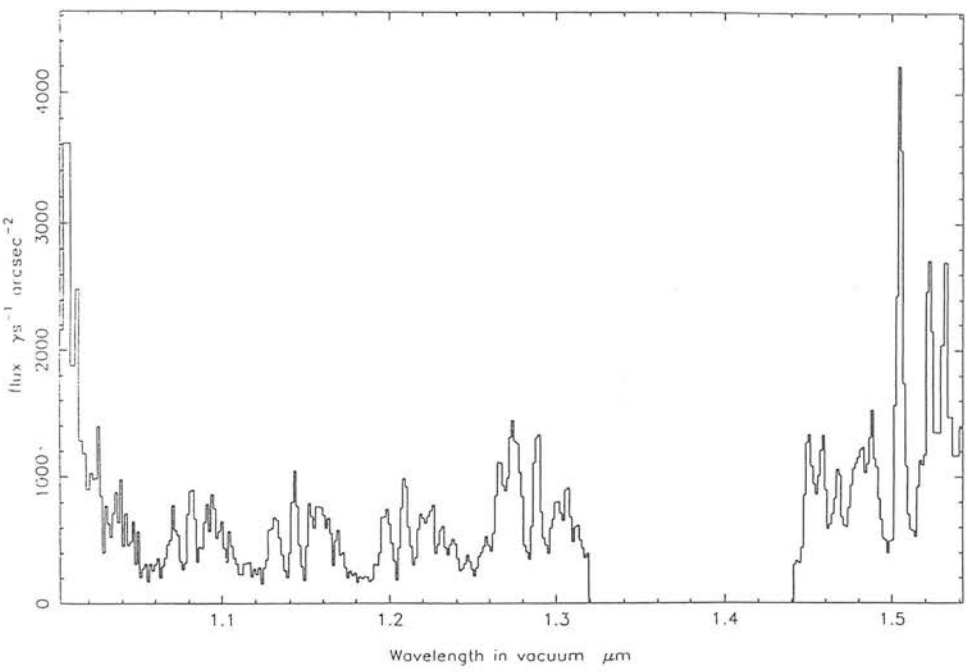


Figure 4.1: The spectrum of OH emission in the J band.

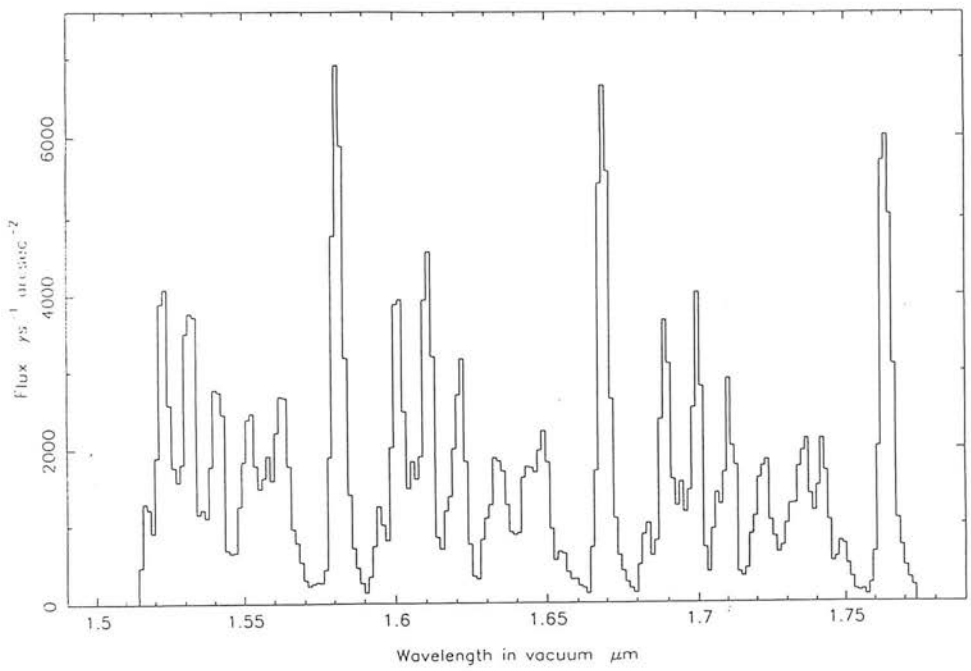


Figure 4.2: The spectrum of OH emission in the H band.

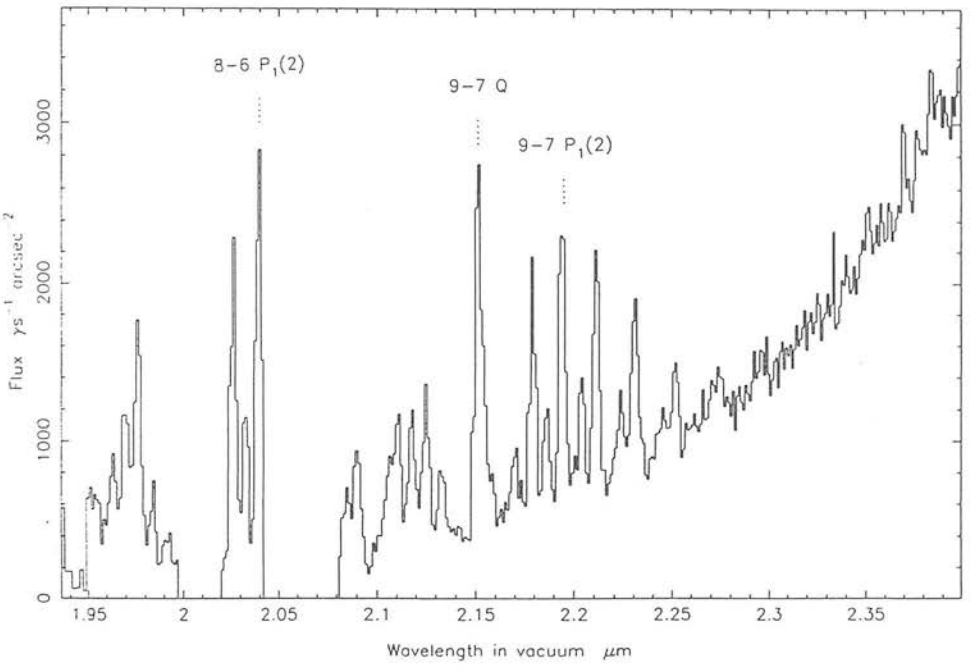


Figure 4.3: The spectrum of OH emission in the K band.

(Figure 4.3); the 9–7 P branch doublets around 2.2μm are resolved.

4.3 Temporal and Spectral Variations

4.3.1 Observations

The information on temporal and spatial variation of the hydroxyl emission was obtained from observations taken with CGS4 in April of 1991. The data on the line strengths of the three lines studied, 8–6 $P_1(2)$, Q(9–7) and 9–7 $P_1(2)$, at 2.0400μm , 2.1518μm and 2.1938μm respectively, were extracted from a set of 30 observations of a high red shift galaxy taken on the night of April 4 between 4.57am and 5.58am, Hawaiian Standard Time (HST), and a further set of 86 observations taken on April 6 between 9.46pm and 1.00am, HST. Our choice of lines enables us to study whether the variations are correlated between vibrational and/or rotational transitions. Each observation consisted of object–sky pairs formed by sequentially offsetting the telescope by $\pm 30''$ along the $3''$ wide slit. Each individual observation was fully sampled by stepping the detector four times over two pixels; the total on–chip exposure time was 80s. The observations were made in NDR mode giving $40e^{-}s^{-1}$ read noise per exposure, so the observations were sky–noise limited. Since the object was not detected in either instance it is assumed that all the line emission at the wavelengths of the OH lines comes from that radical. Each observation has been dark subtracted and flat–fielded. (The dark current in all cases was negligible being $< 5\ e^{-}s^{-1}$). We estimate that the flat–fielding was accomplished with an accuracy of 1 part in 10^3 from the statistical variation of 60 exposures and that any error resulting from this process will be negligible compared with the Poissonian uncertainty in the sky photon flux.

4.3.2 Data Reduction and Results

Spatial Variations

To determine the existence of any spatial variations of the OH emission on a scale of 90 $\widehat{''}$ we extracted a spatial profile at each of the wavelengths of the lines, at 2.0400μm , 2.1518μm and 2.1938μm , each frame having first been flat–fielded. Neither the data for the night of April 4 nor those for April 6 showed any evidence of spatial variation along the slit greater than the Poissonian and flat–fielding errors (1:10^3). As discussed in Section 4.4 this is consistent with the gravity wave models of this emission variation. By considering typical parameters for such a disturbance we find that its effect on the spatial structure of the emission would be dominated by the Poissonian and flat–fielding errors.

Temporal

The spatially flat nature of the OH emission during the observations permits the use of the whole slit to find the signal strength; the values for the OH emission strength presented were arrived at by taking the mean signal strength for 16 pixels in the spatial direction. (Bad pixels were ignored to avoid distortion of the mean.) No attempt has been made to calibrate the data. The data reduction system for CGS4 provides signals in "data numbers per exposure"; to allow us to correctly estimate the Poissonian error, $\sqrt{\text{photoelectrons}}$, we have converted this to electrons by multiplying by the electrons/data number gain for CGS4 (24). During our observations the telescope tracked the object over a range of airmass. The effect of this increasing airmass is to enhance the OH signal strength, as the column density of the emission region is increased. This can be compensated for by dividing the mean signal by the Van Rhijn factor , V (Van Rhijn 1921), at the zenith angle of the observation, where

$$V = (1 - (\frac{R}{R+h})^2 sin^2 z)^{-\frac{1}{2}} \tag{4.1}$$

and R is the radius of the Earth (6378.17km), h is the height of the emitting layer (85km) and z is the zenith angle of the observation. Figures 4.4 and 4.5 show the

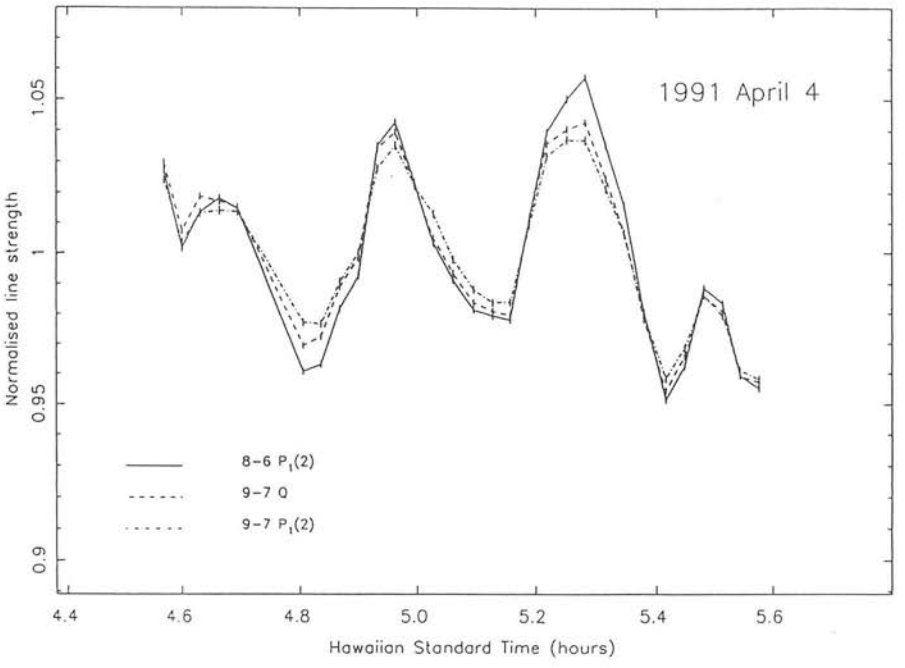


Figure 4.4: Temporal variation of the OH 8–6 $P_1(2)$ line at 2.0400μm , the OH 9–7 Q branch at 2.1518μm and the OH 9–7 $P_1(2)$ lines at 2.1938μm , 1991 April 4th.

temporal variations for the nights of April 4 and 6 respectively; the line strengths have been normalised to the mean line–strength for all the observations and the errors shown on these plots are the Poissonian errors calculated for 16 spectra coadded along the slit.

The data from April 4 show strong evidence of a wave–like structure modifying the emission: the amplitude of the variations is a factor of ten larger than the Poissonian error and is 5% of the total line strength; the period is 640s±40s. By ratioing the line strengths for each combination of pairs of lines we see that the short period, wave–like structure is retained. This effect is most marked in the case of the 8–6 $P_1(2)$ line ratioed with the 9–7 $P_1(2)$ line but is also present less significantly in the other cases, (9–7)Q/(9–7)$P_1(2)$ and (9–7)Q/(8–6)$P_1(2)$ (Figure 4.6).

From Figure 4.6, examination of the temporal behaviour of the emission shows that the lines are varying in phase with each other but that the amplitude of the change in

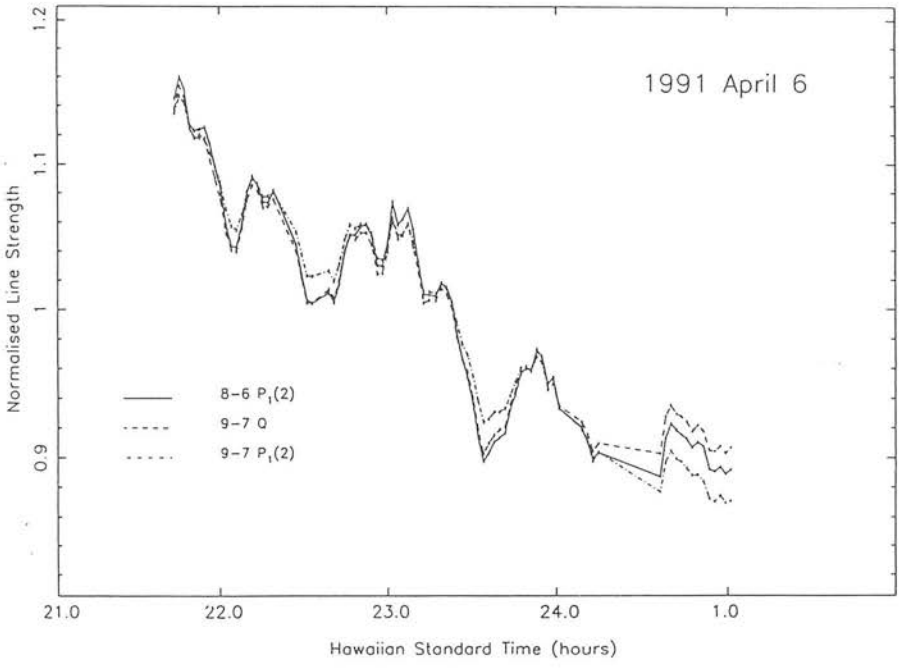


Figure 4.5: Temporal variation of the OH 8–6 $P_1(2)$ line at 2.0400μm , the OH 9–7 Q branch at 2.1518μm and the OH 9–7 $P_1(2)$ lines at 2.1938μm , 1991 April 6th.

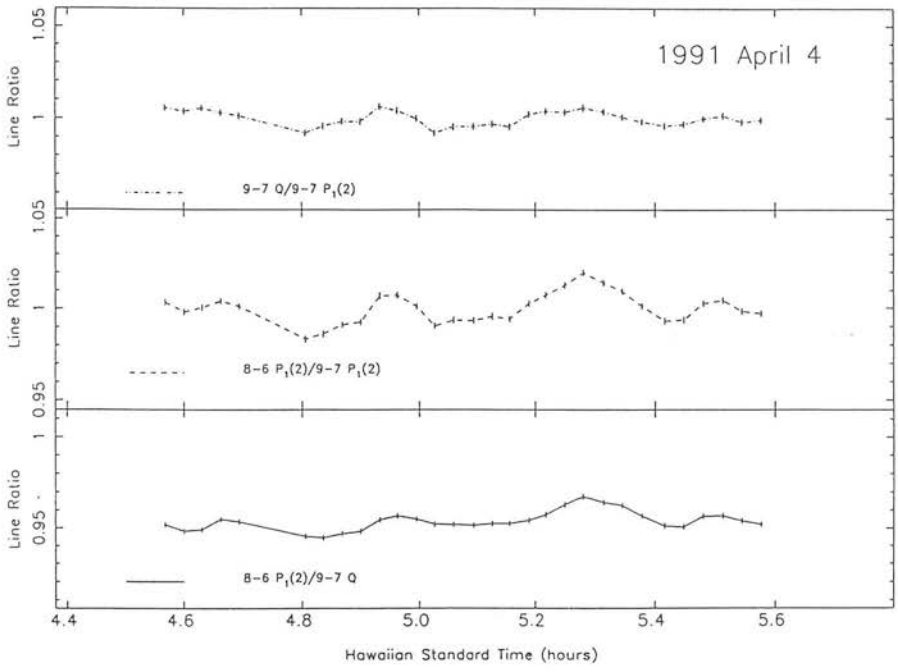


Figure 4.6: The ratio of the fluxes of the 8–6 $P_1(2)$ to the 9–7 $P_1(2)$ line, the 9–7 Q branch to the 9–7 $P_1(2)$ line and the 8–6 $P_1(2)$ line, 1991 April 4th.

emission intensity is not constant for each line, being greatest for the 8–6 $P_1(2)$ line, followed by the (9–7) Q branch and least for the 9–7 $P_1(2)$ line, 10%, 9% and 8% respectively. The data for April 6, extending over a period of 3hrs. also shows wave–like structure superimposed on a global decrease of the emission intensity between 10pm and 1am local time. The amplitudes and periods of the waves are not so regular as for April 4; the range of periods is from about 5mins to 20mins, with the variation in line strength from a few to 10%. The global decrease in the emission strength is 46% $\pm$ 5% over the range of the observations. On this night also, ratioing the emission for pairs of lines shows that the short period wave structure is preserved. Again the (8–6)$P_1(2)$/(9–7)$P_1(2)$ ratio demonstrates this effect most dramatically (Figure 4.7). The effect is not so marked as on the night of April 4, the change in emission intensity for the three lines is 5% for the (8–6)$P_1(2)$ line, 3% for the (9–7)$P_1(2)$ and 4% for the 9–7 Q branch (Figure 4.5).

As well as this simple analysis of the periodic behaviour, Fourier analysis of the data for each line was attempted. The data sets are not particularly suited to this analysis, as few cycles are present during the total observation time. The April 4 data does show one definite peak, with a period of 746s $\pm$ 80s (Figure 4.8), which is in good agreement with the period of 640s$\pm$ 40s estimated above. For the night of April 6, the only significant peaks in the power spectrum were at periods equal to the total observation time used, 5600s, and half that period, 2800s.

4.4 The Excitation of and Variations in OH Emission

As described in Section 4.1 1, the airglow arises mainly from excited vibrational–rotational levels of the OH radical. Bates & Nicolet (1950) determined that the hydroxyl radical is excited chiefly through the reaction

$$H + O_3 \rightarrow OH^* + O_2 \tag{4.2}$$

and this was reinforced by Llewellyn *et al.* (1978). This reaction populates the levels $6 \leq v \leq 9$; lower levels are populated through vibrational relaxation. The emission

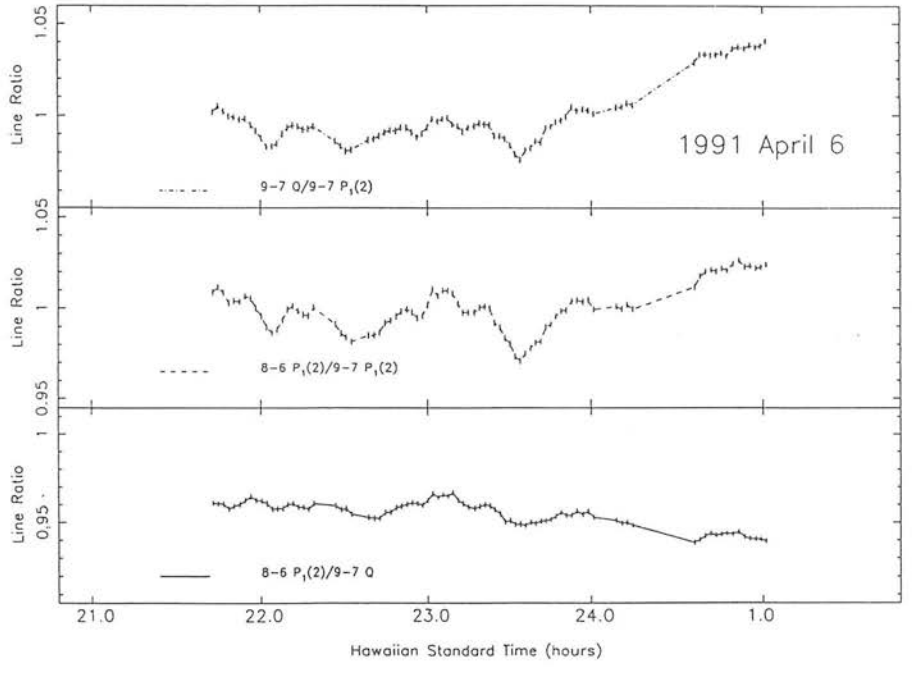


Figure 4.7: The ratio of the fluxes of the 8–6 $P_1(2)$ to the 9–7 $P_1(2)$ line, the 9–7 Q branch to the 9–7 $P_1(2)$ line and the 8–6 $P_1(2)$ line, 1991 April 6th.

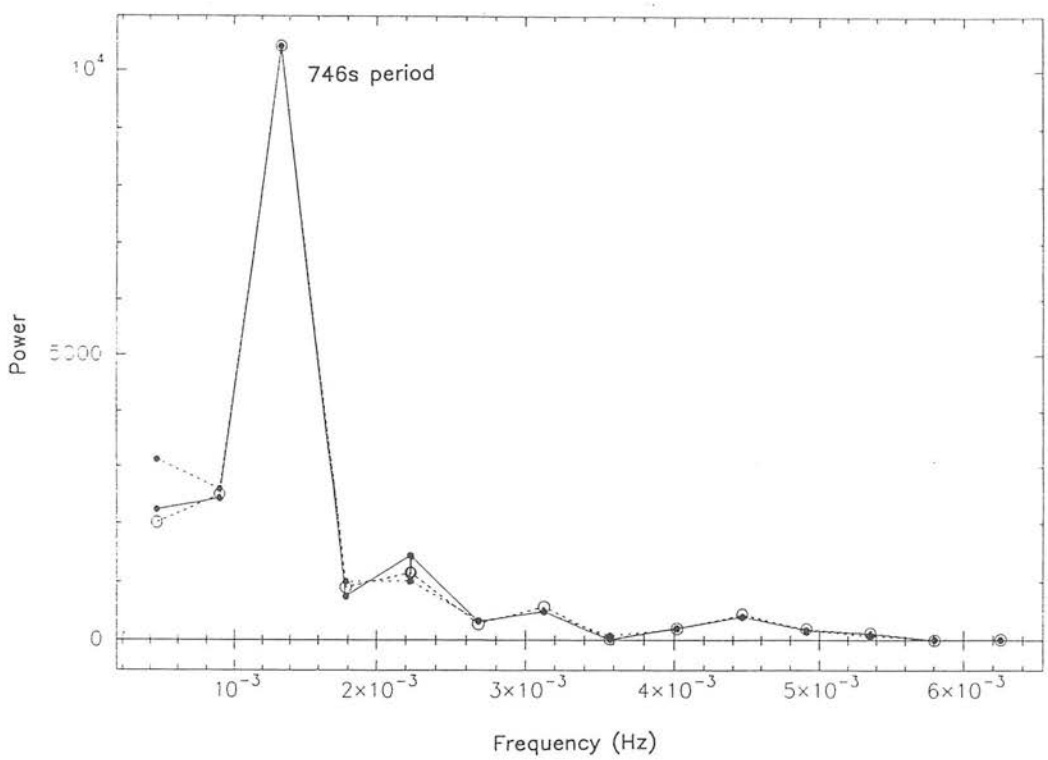


Figure 4.8: Fourier Spectrum of the data from the night of April 4 showing a strong peak for a period of 746s±80s.

occurs mainly through radiative cascade; the OH^- has a radiative rate $A \sim 20s^{-1}$. The depopulation of the excited states is governed by quenching by other atmospheric components, notably

$$OH^* + O \rightarrow H + O_2 \tag{4.3}$$

and by vibrational relaxation (Llewellyn *et al.* 1978). The above quenching reaction is also thought to cause excitation of the oxygen molecule to higher electronic states, which then emit at around 1.27μm forming another component of the non–thermal background emission although this is not the dominant method of excitation of O_2. The correlation between emission from these two species has been reported (see e.g. Llewellyn & Solheim (1978)).

Shorter period variations occur in the nightglow intensity, due to the passage of gravity waves through the ionosphere. The existence of such waves has been well documented and modelled. (Hines 1960; Tuan 1976; Yu*et al.* 1980) These transverse ionospheric disturbances (TIDs) cause local, periodic variation in the temperature and density of the atmospheric gases, which in turn affects the intensity of the OH emission. Hatfield *et al.* (1981), using the gravity wave models of Tuan (1976) and Yu *et al.* (1980), have derived the mechanism for the creation of structure in the OH emission. They consider the effect of the passage of a gravity wave on the unperturbed altitude–number density profiles of hydrogen and ozone derived by Good (1976) and find that a gravity wave of horizontal velocity of $v \leq 33ms^{-1}$ will produce structure in these profiles. The resulting differential in the number density of the key reactants in the chemical excitation of OH, H and O_3 at a given altitude results in an enhancement of the OH emission at that altitude as it is governed by the relation

$$I = \eta \sum k[H][O_3] \tag{4.4}$$

where η is the efficiency of photon emission per OH radical formed and k is the reaction rate, $1.5 \times 10^{-12}T^{1/2}cm^3s^{-1}$ (Nicolet 1971). The nature of the structure is a stationary vertical pattern and a transverse horizontal wave, whose phase velocity is the horizontal speed of the gravity wave. The emission peaks at $80 - 105$ km in altitude, which is

consistent with observations, including the measurements of 85km by Hersé *et al.* (1989) and Taylor et (1987). This kind of structure has been imaged using an infrared detecting TV system (Taylor *et al.* 1987) and photographically (see e.g. Peterson and Kieffaber 1973). From examination of the OH emission structure, Taylor *et al.* determined that the wave had a horizontal wavelength of 26 ± 3 km, a horizontal phase velocity of $38 \pm 2ms^{-1}$ and a period $\tau \sim 11.4$ min for the OH structure, and concluded that the structure is due entirely to the passage of the gravity wave of the same period as the emission, $\tau \sim$ 11.4 min. Allowing for the angle of propagation, the wavelength of the gravity wave is $\lambda = 12.5$km, and its phase velocity $v_{ph} = 18.5ms^{-1}$.

An alternative to this theory of the effect of the passage of a gravity wave on the chemistry of the OH emission is proposed by Moreels and Hersé (1977) to explain their images of airglow structure obtained at Pic du Midi. Envisaging the OH emissive layer as being of constant thickness but deformed by the passage of the gravity wave to resemble a corrugated sheet they calculate an enhancement effect analogous to the Van Rhijn effect (Equation 4.1) which produces the stripes of bright emission common to all images of OH structure.

In this case, no such spatial variation across CGS4's $90\widehat{''}$ slit was detected. However, using the parameters derived from our April 4 data, this non–detection can be explained by simple geometry. Considering the variation of the OH emission to be a sine wave at a height of 85km, taking a wavelength of 25km and a horizontal phase velocity of $40ms^{-1}$, to be consistent with the observations of Taylor *et al.* (1987). Using these parameters with our measured variation in emission intensity of $\sim 10\%$ the spatial variation expected across the $90\widehat{''}$ slit of CGS4 would be 1 part in 7×10^3, a smaller effect than the combined Poissonian and flat–fielding errors (1:10^3). Moreover, our calculations show that this 10% variation in the intensity would be seen on a timescale of ~ 5 minutes, consistent with the results from our April 4 and April 6 data, where the periods range from 5 to 15 minutes. The complexity of the structure seen on the night of April 6 could be due to the presence of several gravity wave systems modifying the emission.

The global decrease during the night is predicted by many models of the night-glow emissions which calculate OH emission strength from the nocturnal changes in the

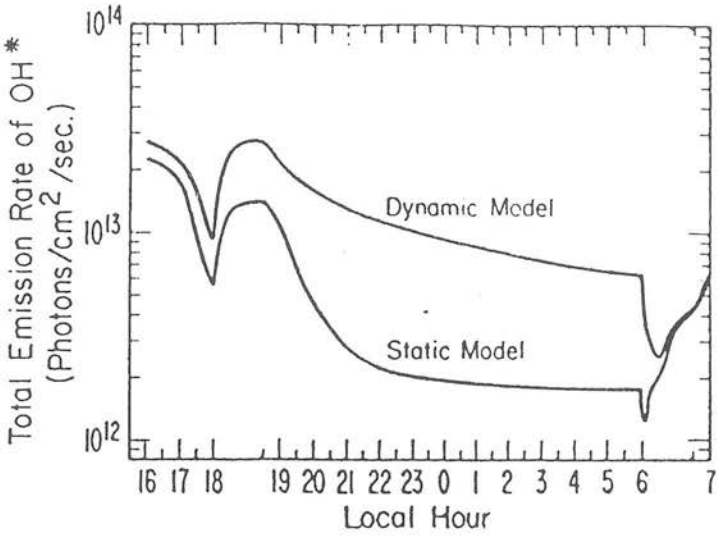


Figure 4.9: Shimazaki & Laird's dynamical model of the upper atmosphere constituents (1970).

altitude/number–density profiles of the minor atmospheric constituents (eg. Gattinger 1969, 1971; Shimazaki & Laird 1970.). The dynamical model of Shimazaki & Laird (1970) (Figure 4.9), which also allows for the effects of vertical diffusion of these constituents, predicts a 50% variation in the OH emission rate between the hours of 10pm and 2am, which is in excellent agreement with our value of 46% $\pm$ 5% for the same period. No such trend was seen in the data for April 4, presumably due to the local time of observation (5am to 6am HST); Shimazaki and Laird (1970) predict that the emission rate is constant at this time.

4.5 RESIDUAL OH SIGNAL IN OBSERVATIONS

To determine an optimum mode for observations of extended objects we have modelled the residual OH signal for an hour of observations. The variation is again considered to be a sine wave with an amplitude of 10% of the OH signal strength as seen in the data of April 4. The hour of observations is made up of object–sky pairs; it is assumed that the end of an object observation and the start of a sky observation are concurrent. For

a given exposure time the residual signal for an object–sky pair is calculated thus:

$$res\_oh_i = \int_{t=s_{obs}}^{e_{obs}} A sin\frac{2\pi t}{\tau} - \int_{t=s_{sky}}^{e_{sky}} A sin\frac{2\pi t}{\tau} \tag{4.5}$$

where res_oh_i is the residual OH signal on one sky/object pair, s_{obs} and s_{sky} are the start times of the object and sky observations with e_{obs} and e_{sky} the end times, A is the amplitude of the disturbance, here 10% and τ is the gravity wave period. This is repeated until the sum of the object and sky exposures is one hour, giving a number of pairs associated with the exposure time. The mean residual for this group of observations is then found as follows:

$$mean\_res\_oh = \frac{1}{n}\sum_{i=1}^{n} res\_oh_i \qquad \text{where} \quad n \sim 3600/(2\tau) \tag{4.6}$$

A further factor which influences this mean residual is the phase of the variation at which the group was started. To allow for this, the mean residual for 500 groups of observations started at random phases was taken and the noise from this mean residual calculated as the standard error of these residuals. To determine the optimum exposure time, we apply the criterion that the noise from the residual OH should be 1% of the photon noise from the OH signal itself and find the exposure times which comply. There are a number of assumptions implicit in this simple model; a more sophisticated model would be required to take into account the fact that the observations become read–noise limited for short exposure times. Also, the residuals for a given phase are assumed to be without error. These assumptions can be justified on the grounds that gross uncertainties in the variation parameters already restrict us to generalisations about the noise behaviour. The results are presented in Figure 4.10, which show residual emission as a function of the exposure time normalised to the period.

4.5.1 Results

The mathematical form of the residual as a function of exposure time is clearly a $sinc^2$ function with maxima and minima occurring as a function of the period of the variation. If we knew the precise value of this period we would be able to select an exposure time

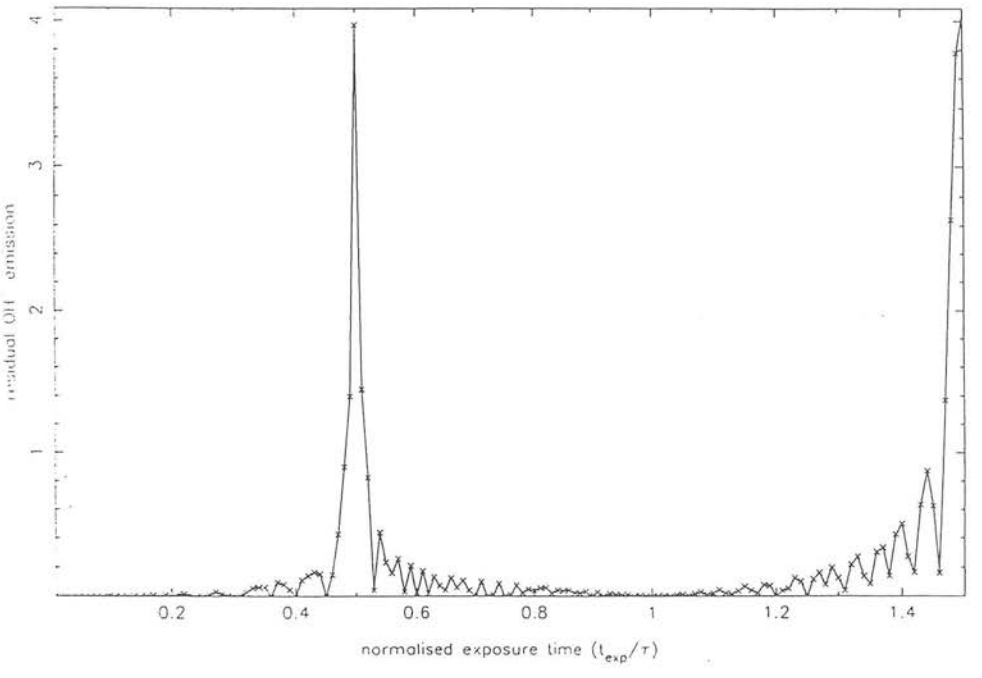


Figure 4.10: The residual OH signal in an hour of observations for a range of exposure times normalised to the period of the emission variation.

such that the residual goes to zero. However, this possibility is not open to the observer. Our calculation of the exposure times for which the residual is 1% of the sky noise show that this condition is fulfilled for ranges of exposure time from 0s to 0.1τ and 0.9τ to 1.1τ, where τ is the period of the variation. At the other extreme, an exposure time of half the period may result in a residual OH signal many times greater than that signal itself, where the maximum, max_oh, is determined thus:

$$max\_oh = \frac{2A\tau}{\pi} \tag{4.7}$$

with A and τ are defined as in (5). Taking a period of 600s, as seen on April 4, to be typical restricts the observer to using only exposures in the 0s to 0.1τ range; the stronger sky emission lines integrated for 600s would saturate array detectors currently in use. So, we are left with the requirement that the on–chip exposure time, before off–setting to the sky, should be $\leq 0.1\tau$, or in the range 30s to 90s for variations in the scale of 5mins to 15mins as seen in our data. It is important to note that this is the simplest model of OH emission; the presence of a number of periods in the data, such as for April 6, would alter these conclusions. Further work in this field, including possibly characterisation of the OH emission for a given site and season will be necessary before we can confidently predict optimum exposure times.

This simple model of the residual OH emission has indicated that the associated noise due to the variations can dominate even the sky–noise in the 1–2.5μm region. The additional noise contribution provides a fundamental limit to the sensitivity achievable with infrared instruments, and is particularly damaging to observations of faint extended sources.

4.6 CONCLUSION

The $1 - 2.5\mu$m spectrum of the atmosphere above Mauna Kea has been presented, along with identification of the OH emission lines, the most significant component of the infrared nightglow. Long–slit spectra from two nights' observations with CGS4 have provided us with information on the temporal and spatial variations of this emission. By

examining the behaviour of three of the OH lines we have shown that the spatial variation is negligible over CGS4's 90$\widehat{\text{m}}$ slit and that the temporal variation over periods of tens of minutes to three hours is consistent with theories of both the short period variation induced by the passage of a gravity wave and of the long term, diurnal variations in emission strength. This variation was observed to have periods in the range 5mins to 15mins, with the variation in line strength lying in the range 3% to 10%. In addition to this, ratioing the line strengths for each of these lines has demonstrated that the emission from individual vibrational–rotational OH lines is not modified identically by the passage of a gravity wave, ie the line ratio between any pair of lines is not constant although the general trend, increasing or decreasing flux, is in phase for each line. Modelling of the additional noise from residual OH emission has shown that restricting on–chip exposure times in the range 30s to 90s is advisable to combat this source of noise. However, these values are presented with the caveat that OH variation parameters are poorly determined and that the amplitude of variation and its period strongly affect the residual and choice of exposure time.

Chapter 5

Pure Fluorescent Emission from Hubble 12

Introduction

In this Chapter, and that which follows, astronomical results obtained with CGS4 are presented. Observation of the emission from molecular hydrogen is one of the areas in which the NIR region of the spectrum is invaluable, and is a frequently employed method for determining the chemical and physical conditions in regions of recent star formation. The mechanisms whereby the H_2 can be excited have been detailed in Section 1.3.2; this Chapter describes an investigation into the nature of the H_2 emission from Hubble 12, which has been carried out in collaboration with Antonio Chrysostomou, Tom Geballe, Peter Brand and Matt Mountain. This work has been accepted for publication in *Monthly Notices of the Royal Astronomical Society.*

Hubble 12 is a young, bright planetary nebula whose ionised core is surrounded by a $4^{\widehat{11}}$–$7^{\widehat{11}}$ torus of H_2 emission (Dinerstein *et al.* 1988). The first NIR spectrum of the H_2 emission from this region was obtained by Dinerstein *et al.* (1988) from which it was deduced that the line ratios indicated fluorescent emission from a gas of density 10^4–$10^5 cm^{-3}$ excited by UV photons from the central source. This spectrum is shown in Figure 5.1.

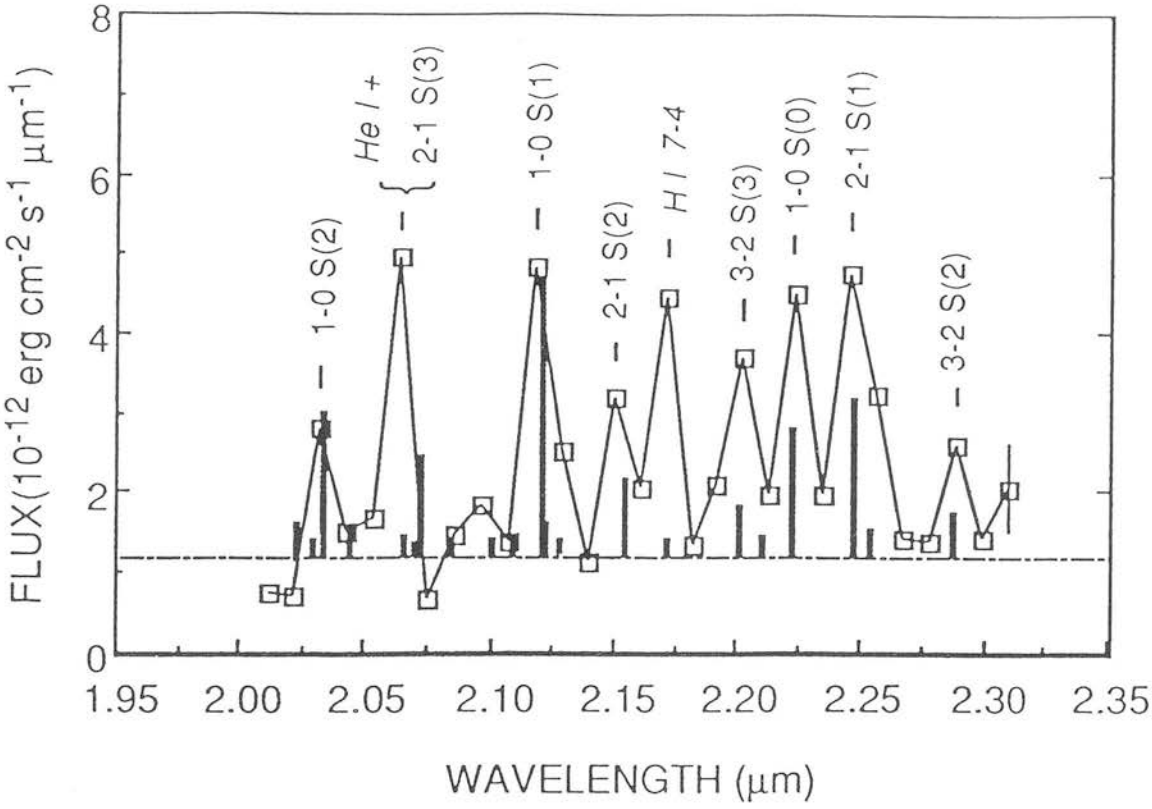


Figure 5.1: The spectrum of Hubble 12 obtained by Dinerstein *et al.*. The unusual appearance of this spectrum is due to fact that the lines are undersampled.

The most recent models of the radiative excitation of H_2 (Sternberg & Dalgarno 1989; Burton, Hollenbach & Tielens 1990) show that the earliest and simplest tests which distinguish between excitation mechanisms are insufficient in many circumstances; these models are outlined in Section 1.3.2. In dense regions, or regions where high density condensations are present as observed in many photodissociation regions (PDRs) (*e.g.* M17 (Stutzki *et al.* 1988, Chrysostomou *et al.* 1992), the Galactic centre molecular ring (Chapter ??, Ramsay, Mountain & Geballe 1993) and the Orion bar (Parmar 1991)), the lower vibrational levels may reach local thermodynamic equilibrium (LTE) through collisions with the gas causing the line ratios and excitation temperatures to appear more thermal than fluorescent. The relative intensities of the lines are also affected by the spin degeneracy of the levels. For a thermal population distribution the ratios of the spin degeneracy for ortho (odd) and para (even) levels is 3:1, but may deviate from this for pure fluorescent or mixed fluorescent and thermal excitation (Tanaka *et al.* 1989).

The Dinerstein *et al.* (1988) spectrum, obtained $4\widehat{''}$ east of the central star, appears to be consistent with pure fluorescent emission (*i.e.* no thermal component), but it is uncertain due to undersampling of the spectrum, low spectral resolution and uncertainties due to noise. We have re–observed the spectrum in order to test more stringently the purity of the fluorescent emission. In doing so, additional lines of H_2 from higher vibrational levels (as high as v=8) have been detected and it is demonstrated that the H_2 line emission is purely fluorescent.

5.1 Observations and Reduction

The K–band spectrum of Hubble 12 is of a position $4\widehat{''}$E of the central star (23h 23m 57.10s, +57 54 32.5 (1950)) close to that observed by Dinerstein *et al.* (1988), with a $60\widehat{''}$E offset position for sky measurements. The K–window from 2.05μm –2.47μm was measured with one grating setting with a resolving power of $\sim$ 330 (0.0067μm /pixel). The spectra were fully sampled, with three samples per resolution element. The wavelength scale was fixed using the H_2 lines observed and is accurate to 0.0006μm or 0.1pixel. The spectra were flat–fielded with the errors on the flat–fields measured to be $\sim$1:10^3. Flux calibration and correction for telluric absorption lines was achieved by

ratioing the spectra with the standard star BS8585 (K = 3.71mag). The spectral shape of the star was removed by multiplying the resulting spectra by a black–body curve at the assumed effective temperature of the star (9230K). The spectra were normalised to exposure times and stellar magnitude. Brackett γ (2.166μm) was present in absorption in the spectrum of BS8585, and so was removed, by fitting the line profile and dividing the stellar spectrum by the resulting function, before ratioing the spectra of Hubble 12 by that of the star. The final spectrum was produced by summing five spectra along the slit giving a $3.1\widehat{''}$ by $15.5\widehat{''}$ beam.

The line fluxes from the Hubble 12 spectra were calculated by fitting a triangular profile, which results from the convolution of the slit and the square pixels, to the line shape, using a least squares fitting method with the line centre and height left as free parameters. FWHM was fixed at 0.0067μm , a value determined by finding the best–fit FWHM for the Brγ line at 2.166μm , which is the strongest line and is not blended with other lines. The value for the FWHM was then used to fit emission lines from 2.0μm to 2.4μm , the change in dispersion with wavelength across the array ($\sim$0.001μm) having a negligible effect on the values for the line fluxes.

5.2 Results and Discussion

The K–band spectrum obtained is shown in Figure 5.2.

In addition to the HI Brγ line at 2.166μm , the HeI singlet 2p–2s and 4s–3p lines at 2.058μm and 2.113μm , a total of 16 H_2 lines can be seen. Several of these are higher excitation lines than were detected by Dinerstein *et al.*. The strongest of the new lines are the 2–1 S(0) line at 2.356μm , the 3–2 S(1) line at 2.386μm and members of the 1–0 Q branch. The 4–3 S(3) line at 2.345μm is partially resolved from the adjacent and stronger 2–1 S(0) line. In addition, the 8–6 O(3) and O(5) lines (at 2.042μm and 2.211μm respectively), arising from energy levels of $\sim$39,000K, are detected as emission shoulders on the wings of other H_2 lines. The 10–8 Q(1) (2.325μm) is marginally detected, placing a 3σ upper limit on this line of 1.14×10^{-8} W m^{-2} sr^{-1}. The wavelengths of additional higher excitation lines fall in the spectral interval covered, but these lines

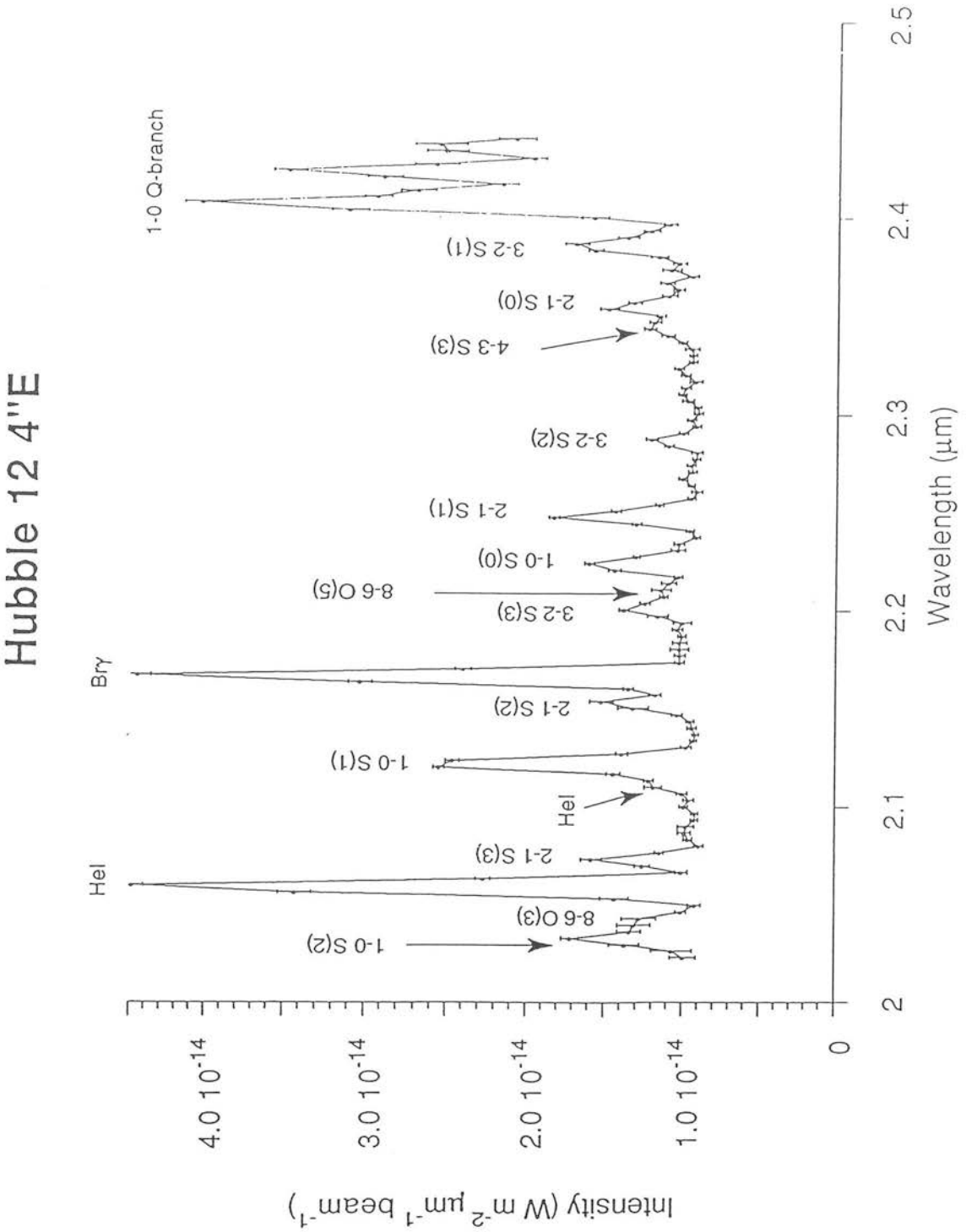


Figure 5.2: K band spectrum of a position 4"E of the central star of the planetary nebula Hubble 12. 16 H_2 emission lines are labelled, as well as emission lines from HI and HeI.

are not detected either because they are coincident with stronger lines or they are too weak to be seen above the noise level. The detected lines and line strengths are listed in Table 5.1. The line strengths for the 1–0 Q–branch could not be determined due to the blending of the lines and the lack of surrounding continuum. From the intensities measured the column densities, excitation temperatures and ortho/para ratios for the different vibrational levels were calculated.

By taking the ratio of two lines with a common vibrational level and for which the rotational levels are either both odd or both even the rotational excitation temperature can be calculated. From Equation 1.2 it can be seen that this ratio can be rearranged to give

$$T_{rot} = \frac{\Delta E}{k \ln \frac{I_i \tilde{\nu}_i A_i}{I_j \tilde{\nu}_j A_j}} \tag{5.1}$$

Suitable pairs of lines were measured for the three lowest vibrationally excited states. For the v=2 level, two pairs of lines were observed giving two rotational temperatures. Both values were the same within the errors. The excitation temperatures for each vibrational level are reported in Table 5.2; the mean value is ~1100K. The vibrational excitation temperature can be calculated from the ratios of lines from a common rotational level. The 1–0 S(1) and 2–1 S(1) lines were used, as they have the highest signal/noise ratio. The vibrational temperature from these lines is found to be ~5000K. An empirical ortho/para ratio was calculated from the observed intensities and rotational excitation temperature by ratioing pairs of lines from adjacent odd and even states and using the equation:

$$\frac{g_{s_{ortho}}}{g_{s_{para}}} = \frac{I_o}{I_p} \frac{A_p (2J_p + 1) \lambda_o}{A_o (2J_o + 1) \lambda_p} e^{\frac{-\Delta E}{kT_{rot}}} \tag{5.2}$$

where $g_{s_{ortho/para}}$ is the spin degeneracy of the odd/even state, $I_{o/p}$ is the observed line intensity, $A_{o/p}$ is the transition probability, $\lambda_{o/p}$ is the wavelength of the line and ΔE is the difference in energy of the upper states. The ortho/para ratios determined for each vibrational level are included in Table 5.2.

For the remainder of the analysis an ortho/para ratio of 1.72±0.17 is adopted, the weighted mean of the measured ratios. Both the high value for the vibrational exci-

Intensities and Column Densities from Hubble 12						
Line	Intensity (10^{-8} W m^{-2}sr^{-1})	degeneracy	wavelength (μm)	A (10^{-7}s^{-1})	Energy Level (K)	$\frac{N_j}{g_j}$ (10^{17}m^{-2})
1-0 S(0)	4.56±0.20	5	2.2233	2.53	6472	50.77±2.2
1-0 S(1)	12.4±0.3	21	2.1218	3.47	6952	22.8±0.5
1-2 S(2)	5.00±0.56	9	2.0338	3.98	7585	18.0±2.0
2-1 S(0)	2.96±0.36	5	2.3556	3.68	12097	24.00±2.9
2-1 S(1)	5.42±0.22	21	2.2477	4.98	12552	7.37±0.30
2-1 S(2)	3.61±0.39	9	2.1542	5.60	13152	9.76±1.05
2-1 S(3)	3.79±0.04	33	2.0735	5.77	13892	2.61±0.03
3-2 S(1)	4.97±0.36	21	2.3865	5.14	17821	6.95±0.50
3-2 S(2)	1.66±0.21	9	2.2871	5.63	18389	4.74±0.60
3-2 S(3)	2.53±0.22	33	2.2014	5.63	19089	1.90±0.17
4-3 S(3)	1.63±0.23	33	2.3445	4.58	23959	1.60±0.23
8-6 O(3)	2.74±0.78	9	2.0418	15.90	38715	2.47±0.70
8-6 O(5)	1.13±0.37	21	2.2107	8.77	39227	0.86±0.28
He I 2p-2s	23.9±0.5		2.058			
He I 4s-3p	2.02±0.24		2.113			
H I Brγ	21.4±0.40		2.116			

Table 5.1: The H_2 , HeI and HI lines present in the K band spectrum of Hubble 12.

Excitation Temperatures and Ortho/Para Ratios		
Vibrational level, v	Rotational Excitation Temperature (K)	Ortho–Para Ratio
1	1072±124	2.11±0.28
2†	1174±208	1.36±0.39
2‡	1291±50	1.42±0.27
3	976±83	2.46±0.70
weighted mean from all levels		1.72±0.17

Table 5.2: The excitation temperatures and ortho–para ratios determined from the K band spectrum of Hubble 12. † – rotational temperature calculated using the 2–1 S(0) and 2–1 S(2) transitions and the ortho–para ratio from the 2–1 S(1) and 2–1 S(2) transitions. ‡ – rotational temperature calculated using the 2–1 S(1) and 2–1 S(3) transitions and the ortho–para ratio from the 2–1 S(2) and 2–1 S(3) transitions.

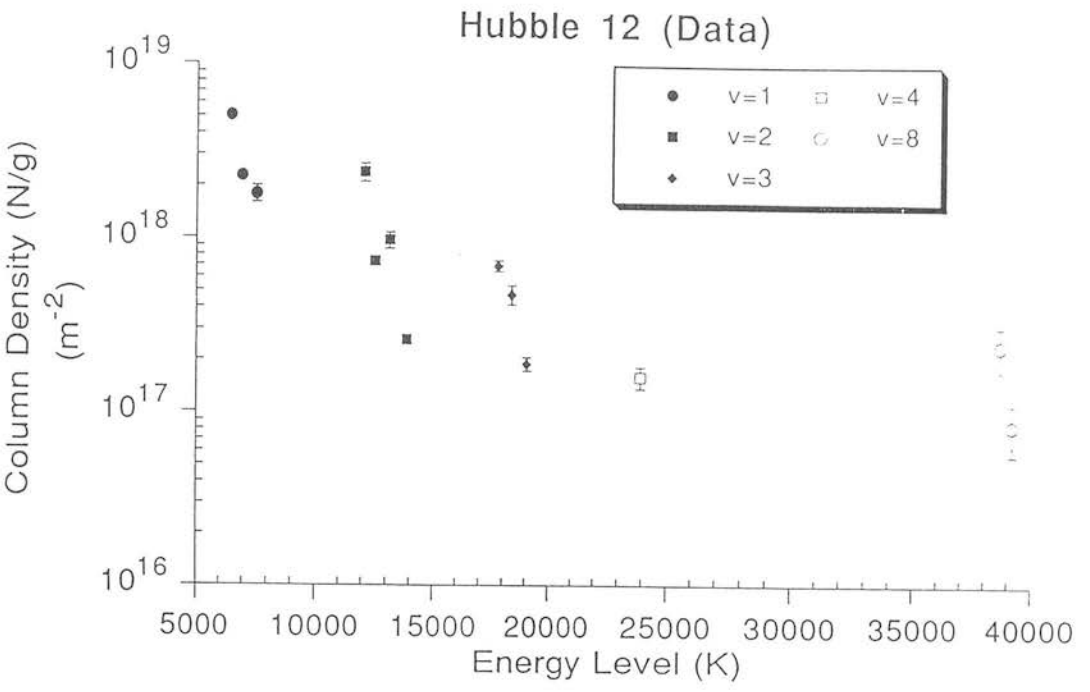


Figure 5.3: Column density against energy level for the 13 H_2 transitions observed in the Hubble 12 K band spectrum. An ortho–para ratio of 3 has been assumed when normalising by the level degeneracy.

tation temperature relative to the rotational excitation temperature and the observed ortho/para ratio of less than 3 are characteristic of radiative excitation, and provide the first pieces of evidence in support of our claim that Hubble 12 is a source of pure fluorescent emission.

Plots of column density against energy level for the transitions observed are shown in Figure 5.3.

Hasegawa *et al.* (1987) show that the appearance of these plots is dependent on excitation mechanism: for thermal excitation, lines from all rotational–vibrational levels will lie along a common curve, for low density UV excitation each vibrational level defines a separate curve along which the rotational levels lie. As increasing gas density collisionally populates the lower–levels of H_2 and the emission spectrum becomes thermalised, the column–density energy curves for the low vibrational levels approach the curve defined by thermal excitation (Burton, Hollenbach & Tielens 1990). Figure 5.3 shows clearly that

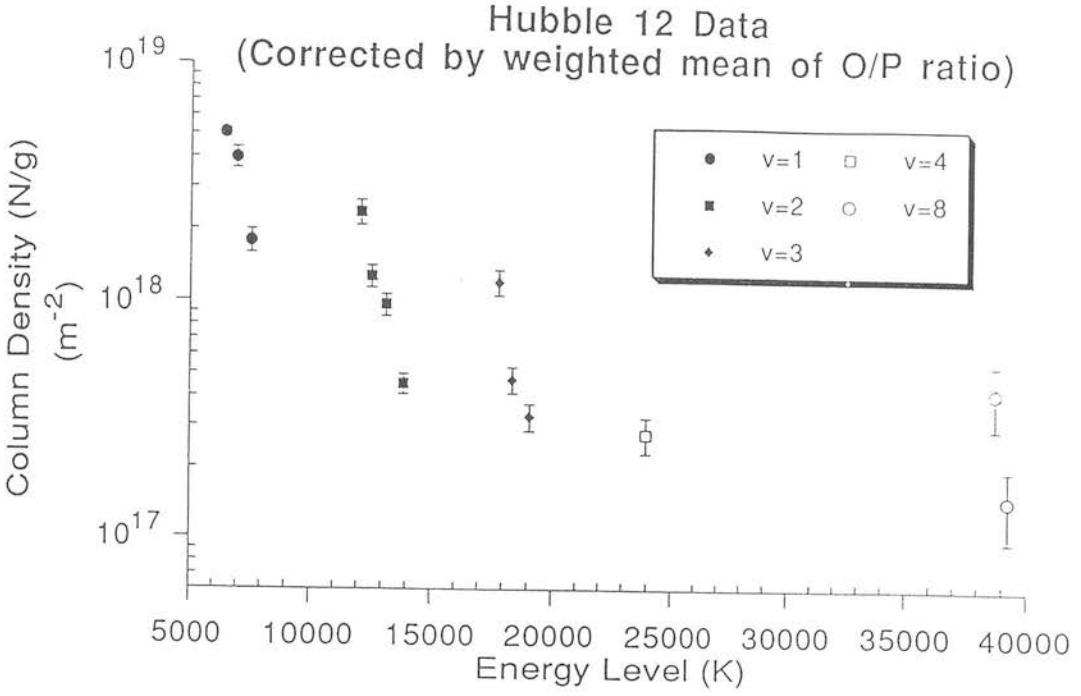


Figure 5.4: Column density against energy level after the column densities have been corrected for the ortho–para ratio having a value of 1.7.

the vibrational levels lie on distinct curves with the rotational levels scattered about the individual curves. This spread of the values is because the ortho/para ratio is not 3, causing the ortho level column densities/degeneracy to be systematically underestimated relative to the para level column densities. Figure 5.4 shows a revised column density/energy level plot after correction for the observed ortho/para ratio; its appearance is as expected for low-density UV excitation. In particular, we see no evidence for thermalisation of even the lowest levels.

In order to confirm that Figure 5.2 shows the spectrum of a low–density gas the column densities after correction for the ortho/para ratio have been ratioed with the predicted column densities for Black & van Dishoeck's Model 14 which is characterised by a density of $3\times10^3 cm^{-3}$ and a UV field of $I_{UV}=10^3 G_o$, where G_o is the UV field of interstellar medium 1.6×10^{-3} ergs $s^{-1}cm^{-2}$ (Tielens & Hollenbach 1985). Model 14 was also corrected for its ortho/para ratio. The results are plotted against energy level in Figure 5.5.

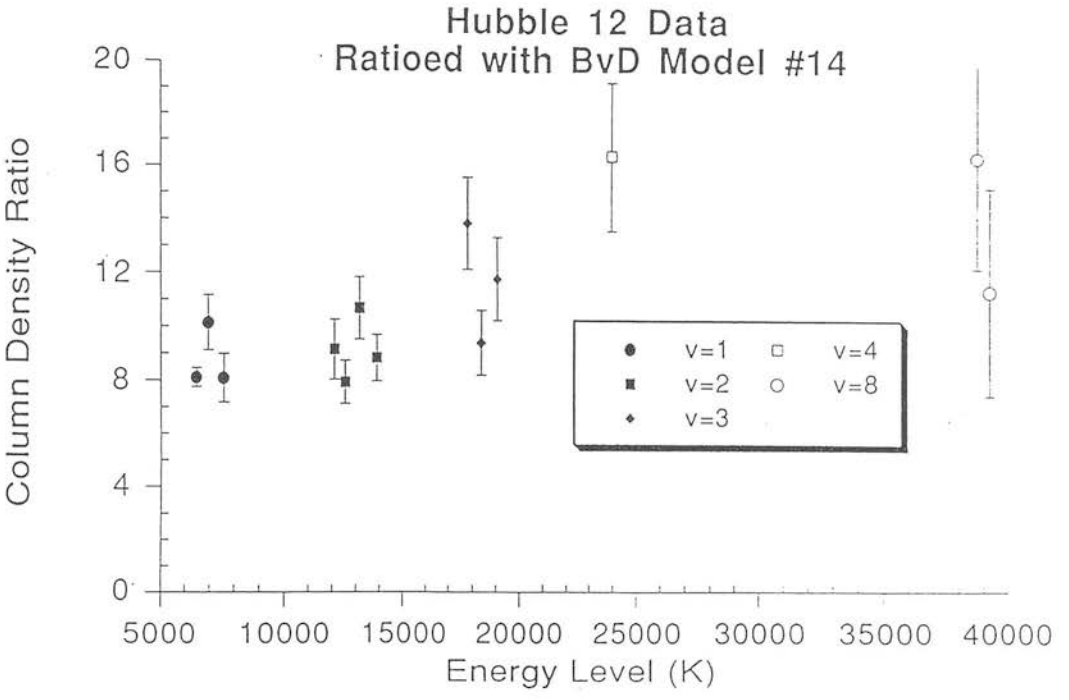


Figure 5.5: The ratio of the column densities of the observed lines with the model values from Black & van Dishoeck's Model 14, which describes pure fluorescent emission from a low-density gas. The horizontal line indicates that there is no evidence for thermalisation of the H_2 round Hubble 12.

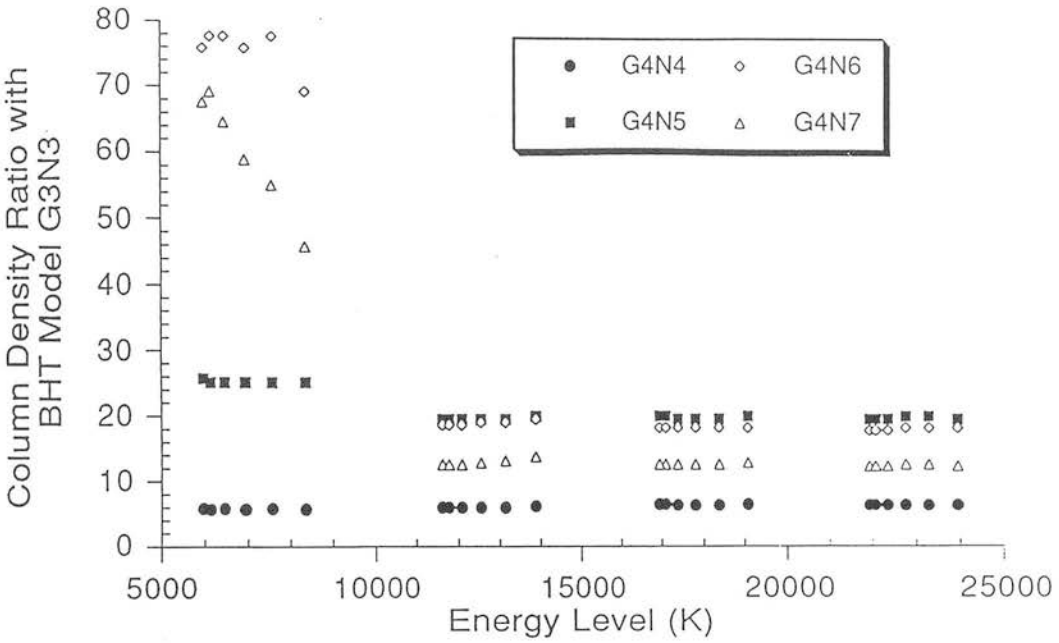


Figure 5.6: The effect of thermalisation on H_2 emission lines is shown in this plot of high density models from Burton, Hollenbach & Tielens (1990) ratioed with Black & van Dishoeck's Model 14,

The points are seen to lie along a straight, horizontal line indicating an excellent agreement between model and observations. In order to illustrate the appearance of the plot if the energy levels were becoming thermalised through collisions in a high–density gas, a similar plot showing the effect of ratioing the lowest density and UV intensity model from Burton, Hollenbach & Tielens (1990) ($n=10^3 cm^{-3}$, $I_{UV}=10^3 G_0$) with models of increasingly higher density gas and higher UV field ($n=10^4$–$10^7 cm^{-3}$, $I_{UV}=10^4 G_0$) is given in Figure 5.6. Increasing the density has the effect of systematically raising the column density of the vibrational levels which are thermalised, notably the v=1 level as dramatically demonstrated by the model results for $n= 10^{6-7} cm^{-3}$. No evidence for trends of this sort is seen in the data.

Therefore the line ratios are consistent with fluorescent emission from a low–density gas exposed to a moderate UV field. It remains to provide the observed intensity within these constraints. Hubble 12 is one of the brightest H_2 sources in the sky: the flux from the

1–0 S(1) lines is 1.24×10^{-7} W m^{-2} sr^{-1}. The UV field from the central star impinging on the molecular cloud at a distance of $4\widehat{''}$ has been estimated to be $3\times10^{4}G_o$ by Dinerstein *et al.* (1988). Comparing the measured flux for the 1–0 S(1) line with the value from the models of Burton, Hollenbach & Tielens (1990) of the appropriate I_{UV}, models with densities of 10^4cm^{-3}, 10^5cm^{-3} and 10^6cm^{-3} will give line intensities for the 1–0 S(1) line of 1.4×10^{-8} W m^{-2} sr^{-1}, 6×10^{-8} W m^{-2} sr^{-1} and 2×10^{-7} W m^{-2}sr^{-1} respectively, where the model intensities are calculated for a line of sight normal to the face of the cloud and for a homogeneous gas. Thus within the uncertainties about the geometry of the source and the effects of a possibly clumpy medium on the observations, both of which would tend to increase the surface brightness, UV excitation of a high density gas can provide the observed intensity . The highest density model can be eliminated, as we would expect to see some evidence of thermalisation with densities of 10^6cm^{-3} (Figure 5.5, Figure 5.6). Thus the observed emission from Hubble 12 can be explained as pure fluorescent emission from a gas of density $\sim10^4$–10^5cm^{-3} illuminated by a UV field, $I_{UV} \sim 10^4G_o$, in agreement with the conclusions of Dinerstein *et al.* (1988).

5.3 Conclusions

We have observed 16 ro–vibrational emission lines from molecular hydrogen in the K–band spectrum of the planetary nebula Hubble 12 taken with CGS4. From the relative intensities of the lines a compelling case can be made for the emission from Hubble 12 at a position $4\widehat{''}$E of the star being pure fluorescent emission, in agreement with the initial results of Dinerstein *et al.* (1988). The observed high vibrational excitation temperature and moderate rotational temperature are indicative of radiatively excited gas, as is the ortho/para ratio which is less than the value of three expected for a thermal population distribution. The relationship between column density and energy levels for the observed lines is as expected for fluorescent emission from a radiatively excited gas. There is no evidence for thermalisation of the lowest vibrational levels, implying a low–density gas. Comparison of the observed line ratios with a model for the emission from a gas of density 3×10^3cm^{-3} and an intensity of UV photons of 1×10^4 (Model 14, Black & van Dishoeck 1987) confirms that the line strengths are purely fluorescent and proves that

no thermalisation of the lower levels is occuring, implying that the emission arises in a low–density ($n \leq 10^5 cm^{-3}$) gas. Finally, the absolute line intensities can be provided by a UV field of 10^4 illuminating a gas of density $\sim 10^4$–$10^5 cm^{-3}$, entirely consistent with the other evidence.

Hubble 12 is clearly an excellent example of a pure–fluorescent emission source and, as one of the brightest objects detected in H_2 , provides an excellent opportunity to study radiative excitation in action.

Chapter 6

H_2 Emission from the Galactic Centre

Introduction

Observations of the the Galactic centre benefit greatly from the application of 2D infrared arrays. Visible light emitted from the galactic centre suffers ~30 magnitudes of extinction, making the central regions invisible to optical instruments and requiring that observations are made at longer wavelengths. The scale of this object makes long–slit spectroscopy an invaluable observational method, whilst the high resolution spectroscopy made possible by the advent of 2D arrays enables the study of the turbulent and complex gas dynamics. In this Chapter, CGS4 observations of the central few parsecs (~150″) of the galactic centre using both the 75l/mm grating and the echelle are presented. These observations have been directed towards resolving some of the questions surrounding the conditions at the Galactic centre through observations of the H_2 ro-vibrational emission from the ring of molecular gas which exists at a distance of 1.5pc from the dynamical centre. The main thrust of the measurements has been to determine the H_2 excitation mechanism, be it radiatively or thermally excited. The high–resolution data have also been used to study the velocity distribution of the gas in the ring. The dynamical behaviour of the gas at the centre, and the implication for the mass distribution, is open to

question. Observations of the molecular ring, the ionised gas and the stellar distribution all provide evidence for either a single point source controlling the dynamics (*e.g.* Harris *et al.* 1985) or an extended mass distribution (*e.g.* Lugten *et al.* 1986).

At the galactic nucleus lie the unusual radio source SgrA* (Lo 1982) and the infrared source(s) IRS16, surrounded by an evacuated cavity in which are seen streamers of low–density, ionised gas (Lo & Claussen 1983) with emission lines up to $\sim$ 700 km s^{-1} broad (Geballe *et al.* 1984). The ionised gas is encircled by a disk of molecular material, extending from 1.5pc–8pc from the centre, which rotates at $\sim$100km s^{-1} (see for example observations of HCN from the ring by Jackson *et al.* 1992). The primary question about the galactic centre concerns the nature of the central source and the way in which the structure and phenomena observed are supported. Two principal theories to explain the excitation and motion of the observed gas find support from the observational evidence: the existence of a mass–loss or in-fall source, possibly a black hole which accretes or ejects matter from the surrounding regions, or the occurrence of a recent burst of star–formation. In the latter case it is anticipated that the H_2 will be radiatively excited by the UV photons from the young stars, whereas a black–hole model would lead to the H_2 being thermally excited in shocks.

The argument in support of the existence of a black–hole as the exotic object powering these phenomena has two mainstays: the proper motion of SgrA* and the dynamics of the ionised gas. A number of workers find SgrA* to be closely associated with the dynamical centre of the galaxy: Serabyn *et al.* (1985, 1988) find the dynamical centre to lie within $5''$ of SgrA* making it an obvious choice for the nucleus. Its status as a black–hole candidate is maintained by the fact that it is an unusual radio object (predicted to be associated with black holes by Lynden–Bell & Rees (1971)) and by studies by Backer & Sramek (1982, 1987) which show its proper motion to be consistent with zero, implying that SgrA* is extremely massive. Observations of the motion of the atomic and molecular gas surrounding the centre imply a mass of $\sim 10^6$Msolar for an object which is unresolved for any of the current observations, placing an upper limit of 1.2×10^{14} m on the size of SgrA*. Countering this is the fact that other phenomena associated with black–holes are conspicuously absent, *e.g.* the existence of sources of gamma and X–rays and the presence of a unique infrared source associated with the centre. No component of IRS16

appeared to be coincident with SgrA* until recent observations at ESO succeeded in resolving the IRS16 complex into many separate components, one of which may be the long sought IR companion to this source. The spatial distribution of these sources is used as an argument against the presence of a massive central object by Saunders (1992), who maintains that the probability of objects being so deeply within the gravitational field of a black–hole is 10^{-4}; this point has been made previously by Forrest *et al.* (1987). The broad line widths of the ionised gas ($\sim$700km s^{-1} within 0.4pc of SgrA* (Geballe *et al.* 1984)) require a massive, compact source if they are to be interpreted as being in bound orbits around the centre. However, if the streamers are taken to be in radial motion in to or out from the central source then winds from IRS16 can provide such line widths.

Direct evidence of a recent burst of star–formation has come from the discovery of a cluster of broad line emission helium stars by Krabbe *et al.* (1991), confirmed by Burton and Allen (1991). The work of Sellgren *et al.* on the stellar population at the galactic centre has shown that the presence of many supergiants and IRS7 imply that a burst of star–formation must have occurred within 100Myr.

A preliminary report of this work, which has been carried out in collaboration with Matt Mountain and Tom Geballe, will appear in the proceedings of the Astronomical Society of the Pacific Conference "Astronomical Infrared Spectroscopy" (1992).

6.1 Observations and Preliminary Reduction

The observations were made with CGS4 on 8th June 1991, 25th June 1991 and 26th June 1991. Observations were made at both high and low resolution during the three nights of observations; the instrumental configurations are described in Section 6.1.1.

Beam switching, or "nodding", was used to remove the background emission from the sky; the telescope was offset $2\widehat{''}$ East–West in all cases. Standard stars were observed to provide flux calibration and to remove telluric absorption lines and the broad-band instrumental response. Flat–field frames taken to remove the spectral pixel–pixel gain variations were accurate to 1:10^4 and each observation was flat-fielded by the online data–reduction facility. All observations of the Galactic centre were made in NDR mode;

typical read–noise measured from dark frames taken during the observing period was 100-150 e^-, and the dark current $5e^- s^{-1}$.

The standard stars used were SAO186838 for the K and H spectra and SAO816328 for the echelle spectra. Early–type stars were chosen so that the spectra are relatively free of absorption features from the stellar atmosphere. Brγ was seen in the standard SAO186838, but was removed by using a least squares algorithm to fit a gaussian profile to the line and the surrounding continuum. The uncertainty introduced in the flux calibration by removing the Brγ is estimated to be 0.2%; relative line fluxes are not affected. The stellar spectrum was then divided by the resulting function to remove the absorption line. Individual spectra were divided by the standard star spectrum to remove the telluric absorption features and the broad-band instrument response; the gross spectral shape of the star was removed by multiplying by a normalised black-body spectrum with the same effective temperature as the star. Flux calibration was achieved by comparing the flux from the object with the known flux from the UKIRT standard observed at comparable airmass.

6.1.1 Low resolution spectra

Low dispersion spectra of the Galactic centre were observed in both the H and K windows using the 75 l/mm grating. The K-band spectra run from 2.03μm –2.44μm with a resolving power, R ($=\frac{\lambda}{\delta\lambda}$) of 333, equivalent to a dispersion of 0.0065μm per pixel. To sample the lines fully the detector was stepped four times over two pixels. Such spectra were taken at five spatial positions around the molecular ring. The 3 $\widehat{''}$ pixels correspond to a spatial scale of 0.1pc at the distance of the Galactic centre (7.7kpc, Reid (1990)). The positions observed (specified as the position of the 29th pixel in the spatial direction) are logged in Table 6.1. Figure 6.1 is a schematic diagram of the slit positions on the H_2 map of Gatley *et al.* (1986).

The H-band spectra, 1.56μm –1.76μm , have R=485 giving a dispersion of 0.0033μm per pixel. These spectra were also oversampled by stepping the detector four times over two pixels. Only the NE and SW lobes of emission were observed. Wavelength calibration was achieved using the OH sky lines with wavelengths taken from Oliva and

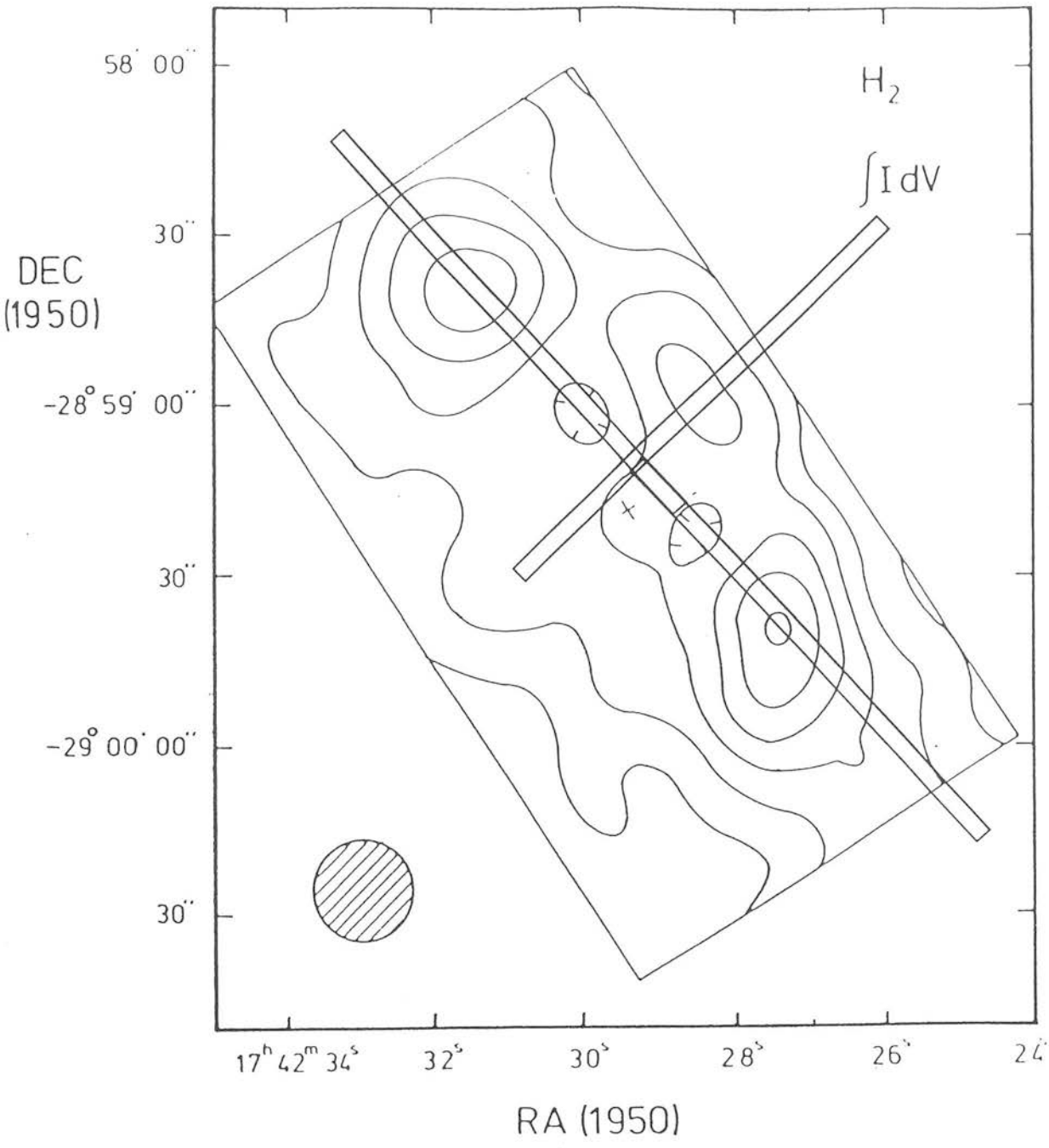


Figure 6.1: The observed slit positions at the galactic centre superimposed on the H_2 map of Gatley *et al.* (1986), from which the positions were selected.

Observing Log					
Object name	Position		Grating μm	Slit PA ° W of N	Time on Source
	α (1950)	δ (1950)			
SW lobe	17h 42m 27s	-28° 59′ 43″	2.215	-43°	14× 60s
			1.66		20× 60s
			ech.		5×60s
NE lobe	17h 42m 31s	-28° 58′ 45″	2.215	-43°	14× 60s
			1.66		13 ×60s
			ech.		6× 60s
NW lobe	17h 42m 28.3s	-28° 58′ 58.5″	2.215	-43°	19× 60s
			ech.		9× 60s
IRS 16	17h 42m 29.3s	-28° 59′ 17″	2.215	-43°	13× 60s
IRS 16 47°	17h 42m 29.3s	-28° 59′ 17″	2.215	47°	11× 60s

Table 6.1: Observing for the Galactic Centre H_2 observations

Origlia (1991). The lack of OH lines above 2.3μm makes wavelength calibration in the K-band spectra unreliable at these wavelengths. The final wavelength scale was fixed using the HI and H_2 lines, which were unambiguously identified using the wavelength scale estimated from the dispersion relation. The H-band spectra do not suffer from this problem, as there are OH lines throughout; the accuracy of calibration in this window is estimated to be 3Å rms.

Line fluxes from the resulting spectra were found by fitting a triangular profile to the observed lines. A correction is made for the extinction to the Galactic centre which is taken to be A_K=3.3 (Wade *et al.* 1987), with a assumed power law dependence of $\lambda^{-1.75}$ (Draine 1988).

6.1.2 High dispersion spectra

Echelle spectra were taken of the NE, SW and NW lobes. The grating wavelength was set to 2.122um, the wavelength of the 1-0 S(1) H_2 line. The theoretical resolving power for this wavelength in 26th order is 8,075 corresponding to a velocity dispersion of 37km s^{-1} per pixel. The true resolution was determined by measuring the width of the two brightest OH emission lines, at 2.125μm and 2.1177μm , which were found to be identical within the $\sim$1km s^{-1} errors, consistent with the non–uniformity of the dispersion calculated from the grating equation; for our configuration (λ=2.122μm , n=26) the change in dispersion across the array is $\sim$1.5km s^{-1} . The mean line width is 37.9km s^{-1} $\pm$1km s^{-1} . Using the echelle grating reduces the slit length to 75$\widehat{\pi}$, with 3.8$\widehat{\pi}$ per pixel in the spatial direction and 2.5 $\widehat{\pi}$ per pixel in the dispersion direction. The slit width is matched to one pixel. Once again the wavelength scale was fixed using the OH emission lines. At this resolution there are only four OH lines present in the spectrum so we use assume a linear fit for the dispersion curve across the array. The dominant source of error is in the accuracy to which the OH emission line wavelengths have been measured, $\pm$15km s^{-1} (Oliva and Origlia 1991). The error on line wavelengths derived after wavelength calibrating is reduced to $\sim$8km s^{-1} , since the accuracy of the fit is greater than the accuracy of the individual lines. This is combined with the error in measuring the line width to determine the error on the gas velocity.

To determine the line widths and fluxes for the resolved lines observed it was assumed that the resulting profile from the convolution of the emission line and the triangular CGS4 instrument profile is a gaussian. The best single gaussian profile for each line was then found using a least squares method with the width, height and central wavelength as free parameters. The profile was the sum of more than one gaussian if a number of velocity components were present. It was estimated that the uncertainty implicit in this assumption was dominated by the error in the wavelength calibration.

6.2 Results and Discussion

6.2.1 Low dispersion

In this Section, and those following, all positions will be quoted as offsets from SgrA*, $17^h\ 42^m\ 29.3^s$, -28° 59′ 18.5″ (Tollestrup *et al.* 1989).

Examples of the spectra obtained are shown in Figure 6.2 and Figure 6.3. The spectrum in Figure 6.2 is of a position offset 5″W, 3.6″N of SgrA* (a radial distance of 0.24pc) and shows emission lines characteristic of the ionised gas in the cavity: the HeI 2p–2s line at 2.058μm and the HI 7-4 Brγ line at 2.166μm . Also present in this spectrum and in the spectrum of the position 1.7″W, 1.3″N is the unidentified line at 2.217μm , discovered by Allen, Hyland & Hillier (1990). This emission, thought to be from either the [FeXII] $^2D_{\frac{5}{2}} - ^2D_{\frac{3}{2}}$ line at 2.205μm or the [FeIII] $^3G_5 - ^3H_6$ line at 2.2178μm , has been identified as arising from a bubble of hot gas formed by a wind from within 2″ of SgrA* (Eckhart *et al.* 1992).

Figure 6.3 shows a spectrum from the molecular ring, offset 28.6″E, 40.4″N (at a radius of 1.92pc). The H_2 emission lines seen are 1-0 S(2) at 2.0337μm , 1–0 S(1) at 2.1222μm , 1–0 S(1) at 2.222μm , 2–1 S(1) at 2.2477μm and the Q–branch lines; the latter are partially resolved. Determination of the H_2 line fluxes, and hence ratios, is rendered impossible in the majority of our spectra by the deep Na and Ca absorption features at 2.208μm and 2.263μm respectively and the CO bandhead around 2.3μm , all from late-type stars at the Galactic centre. The measured lines come only from

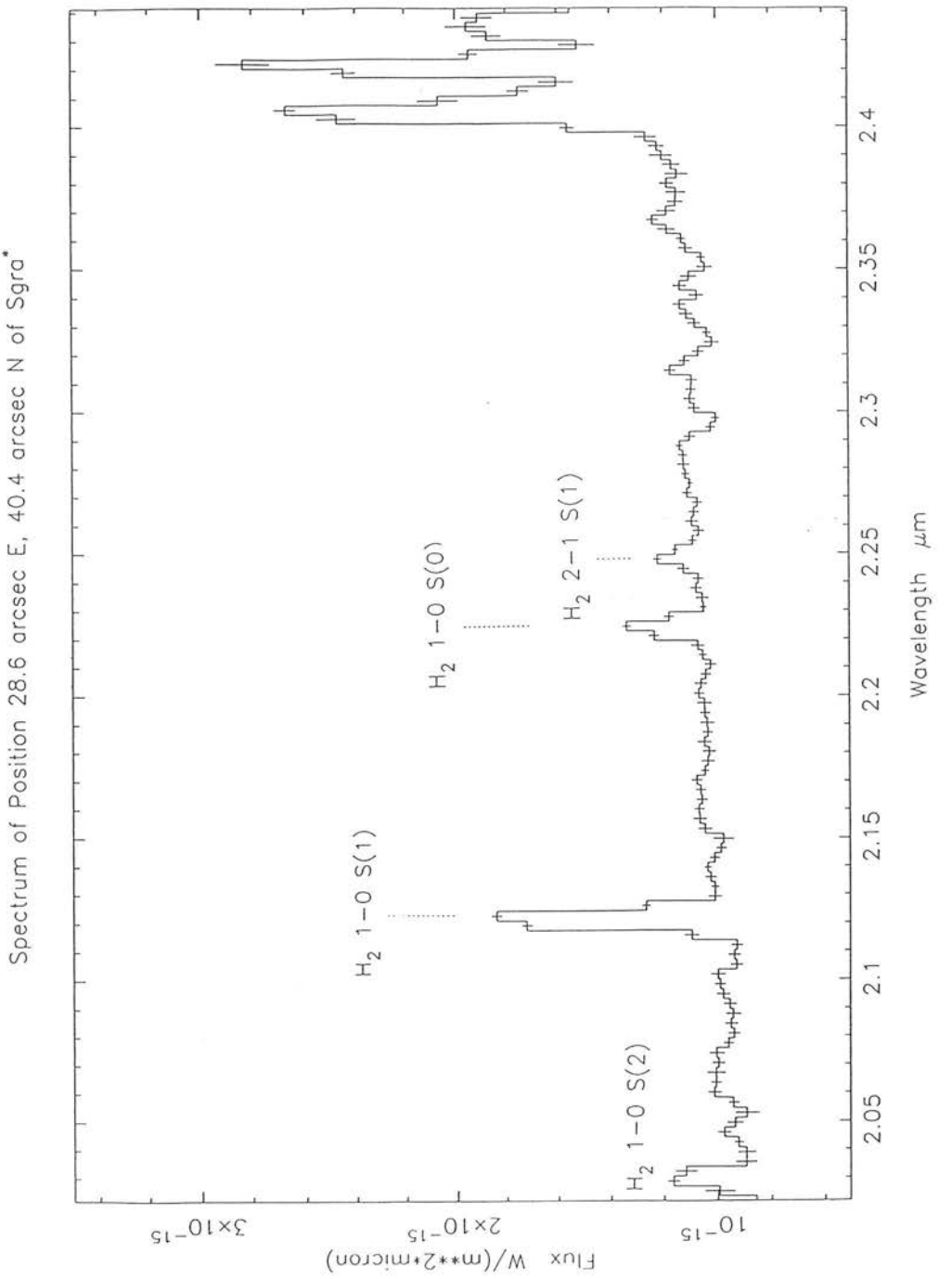


Figure 6.2: CGS4 K window spectrum of a position 5″W, 3.6″N showing emission lines from HI, HeI and the unidentified line at 2.217μm

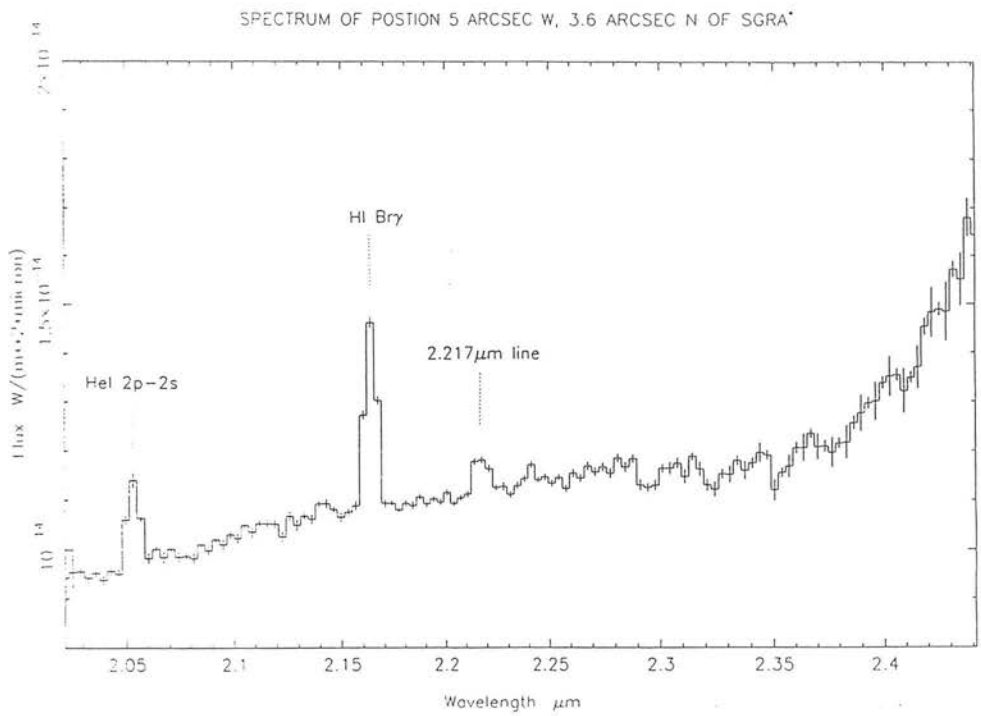


Figure 6.3: CGS4 spectrum of the K–window emission from H_2 at a position in the molecular ring, 28.6″E, 40.4″N. The H_2 emission lines are labelled.

those spectra that do not show strong CO absorption features; the correlation between the presence of the CO absorption and the Ca and Na absorption in the atmospheres of giant stars allows us to predict that there will be no Ca or Na absorption if the stronger CO features are not observed.

Measurable detections of the 1–0 S(1) and 1–0 S(0) lines were made at five spatial positions. Only for the spectrum shown in Figure 6.3 could any other H_2 line strengths be determined reliably; the 1–0 S(2) and 2–1 S(1) lines are seen in addition to the 1–0 S(0) and 1–0 S(1) lines. The positions at which the measurements were made are listed in Table 6.2, along with the line strengths, dereddened column densities and dereddened line ratios. The line strengths of our 1-0 S(1) lines are entirely consistent with previous measurements (Gatley *et al.* 1984, Tanaka *et al.* 1989). Most recently, results from Burton & Allen (1993) give a de-reddened flux for 1–0 S(1) of 1.6×10^{-17} Wm^{-2} in an

$8.46\widehat{''}$ beam for an area NE of SgrA*, compared with our value of $0.2\text{–}1.4\times10^{-16}$ Wm^{-2} in a $9.61\widehat{''}$ beam for a number of different positions throughout the ring. Burton & Allen adopted a value for the extinction of 2mags at 2.1μm , calculate from their 1–0 S(1) and 1–0 Q(3) line strengths. Correcting our data with this value, rather than the value of $A_{2.1}$ of 2.69 derived from the Wade *et al.* result, gives a range of fluxes from 2×10^{-17} to 1×10^{-16} Wm^{-2} for the different spatial positions of our data, in good agreement with Burton & Allen. Uncertainty in the extinction value to be adopted will be a large source of error on this measurement, not only because of the differences in reported values, but also since the extinction at the Galactic centre is known to be patchy (Lebofsky & Rieke 1987). I continue to quote and use the Wade *et al.* value, as the 1–0 Q(3) line often suffers from absorption due to telluric lines. The corrected value for the 1-0 S(1)/2-1 S(1) ratio is 9.23±0.86; the error quoted is derived from the error in the determination of the line flux. An additional error of $\sim \pm0.5$ is expected due to uncertainties in the extinction. The implications of this result for the excitation mechanism are discussed in Section 6.3 below. This value is higher than the value of ~6 quoted by Burton & Allen, at least partly due to their lower extinction value, which would give a 1–0 S(1)/2–1 S(1) ratio of 8.4±1 for our data.

The ortho/para ratio and excitation temperatures were calculated for the spectrum shown in Figure 6.3; a full discussion of these calculations is given in Chapter 5. Using the 1–0 S(2) and 1–0 S(0) lines, the rotational excitation temperature was found to be 1400K, whilst the 1–0 S(1) and 2–1 S(1) lines give a value of 2220K for the vibrational excitation temperature. The ortho/para ratio was calculated as 3.04 using the 1–0 S(1) and 1–0 S(0) lines.

No H_2 lines were observed in the H-band spectra; the HI Brackett series lines were marginally detected in the ionised region; this is discussed in Section 6.3.

These spectra also provide a 2D picture of the spatial relationship of the ionised and molecular gas about the galactic centre as either the HI and/or the 1-0 S(1) H_2 lines were observed in the majority of spectra. Their distribution confirms the existence of the ring of molecular gas encircling the ionised gas (Figure 6.4). From this plot, a measurement of the radius of the ring gives 1.56pc, in excellent agreement with the models of *e.g.*

Galactic Centre Emission Line Data							
position		emission line	Line flux	de-reddened column density	de-reddened line ratio		
α (1950)	δ (1950)		$\times 10^{-18}$ Wm^{-2}	$\times 10^{15}$ cm^{-2}	$\frac{1-0S(1)}{1-0S(0)}$	$\frac{1-0S(1)}{2-1S(1)}$	$\frac{1-0S(1)}{1-0S(2)}$
17h 42m 27.48s	-28° 59' 36.1"	1-0 S(1)	63.17±0.47	20.47±1.53	4.02±0.46		
		1-0 S(0)	1.93±0.16	7.33±0.61			
17h 42m 27s	-28° 59' 43"	1-0 S(1)	4.65±0.82	15.46±2.73	4.60±1.13		
		1-0 S(0)	1.27±0.22	4.82±0.84			
17h 42m 31.16s	-28° 58' 42.7"	1-0 S(1)	3.24±0.98	10.77±3.26	3.10±1.18		
		1-0 S(0)	1.32±0.31	5.01±1.18			
17h 42m 28.6s	-28° 58' 38.1"	1-0 S(1)	6.97±0.20	23.18±0.66	4.34±0.74	9.23±0.86	3.12±0.39
		1-0 S(0)	2.02±0.11	7.67±0.42			
		2-1 S(1)	1.00±0.09	1.86±0.17			
		1-0 S(2)	1.78±0.22	6.24±0.77			
17h 42m 32.5 s	-28° 58' 24.3"	1-0 S(1)	2.39±0.38	7.95±1.26	2.88±0.57		
		1-0 S(0)	1.02±0.14	3.88±0.53			

Table 6.2: Line strengths, line ratios and de–reddened column densities for the observations of the Galactic centre.

Harris *et al.* (1985) and Sutton *et al.* (1990), Jackson *et al.* 1992. A striking feature of these data is the existence of H_2 emission seen within the radius of the molecular ring. A clump of enhanced H_2 and HI emission is present in the spectra at 7.6$\widehat{'}$E, 17.4$\widehat{'}$N and 9.7$\widehat{'}$E, 19.7$\widehat{'}$N ie at a projected radius of ~0.8pc from SgrA*, implying either that there exists H_2 inside the ionised cavity or that we are observing foreground material. To the NW of SgrA* a small enhancement of the H_2 emission is also seen at 7.1$\widehat{'}$W, 1.3$\widehat{'}$N, at a radial distance of only 0.28pc. This intriguing result is repeated in our high resolution spectra (Section 6.2.2) and is discussed in Section 6.3.

6.2.2 Echelle Spectra

The velocities measured for the individual spectra are given in Table 6.3 and line widths in Table 6.5; an example spectrum from a position 33$\widehat{'}$S, 38$\widehat{'}$W of SgrA* at a radius of 2pc, shows two distinct velocity components, one stationary, the other blue–shifted at 30km s^{-1} (Figure 6.5). The line widths are ~50km s^{-1} , clearly greater than the instrumental resolution of 38km s^{-1} , which may be attributable to several causes and which we can use to shed light on the possible excitation mechanisms. Radiatively excited gas will have line widths of ~3km s^{-1} , whereas emission lines excited by a shock front will be broadened (Burton *et al.* 1990). Since the ring is rotating there will be a finite distribution of velocities in the beam of one pixel which can be modelled and may provide sufficient dispersion to explain the observed widths. This is discussed in Section 6.3. Alternatively, the velocity dispersion could be provided by a large number of clumps of different velocities, either shock–excited or excited by UV photons, but which are spatially unresolved.

Position–velocity diagrams of the NE, SW and NW lobes are shown in Figures 6.6, 6.7 and 6.8. H_2 emission is seen throughout the area observed, which extends from SgrA* to a radius of 3.5pc to the NE and 2.4pc to the SW. These data completely support the result from the low–dispersion spectra that there is molecular line emission arising from within the central cavity, as the echelle spectra show emission to within 0.4pc of SgrA*. The strongest emission, which appears as a distinctly separate clump of emission in the P–V diagram of the SW lobe (Figure 6.7), lies between α = 11.8$\widehat{'}$W–6.6$\widehat{'}$W, δ =

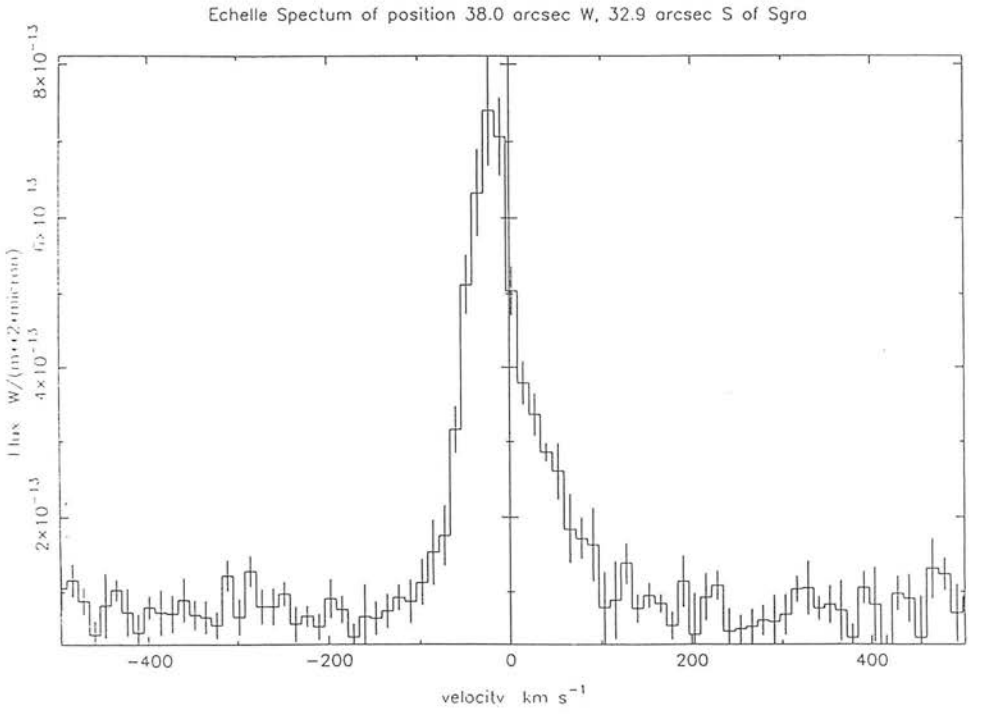


Figure 6.5: CGS4 Echelle spectrum of a position $33\widehat{''}$S, $38\widehat{''}$W of SgrA* showing the velocity structure present in many of the spectra.

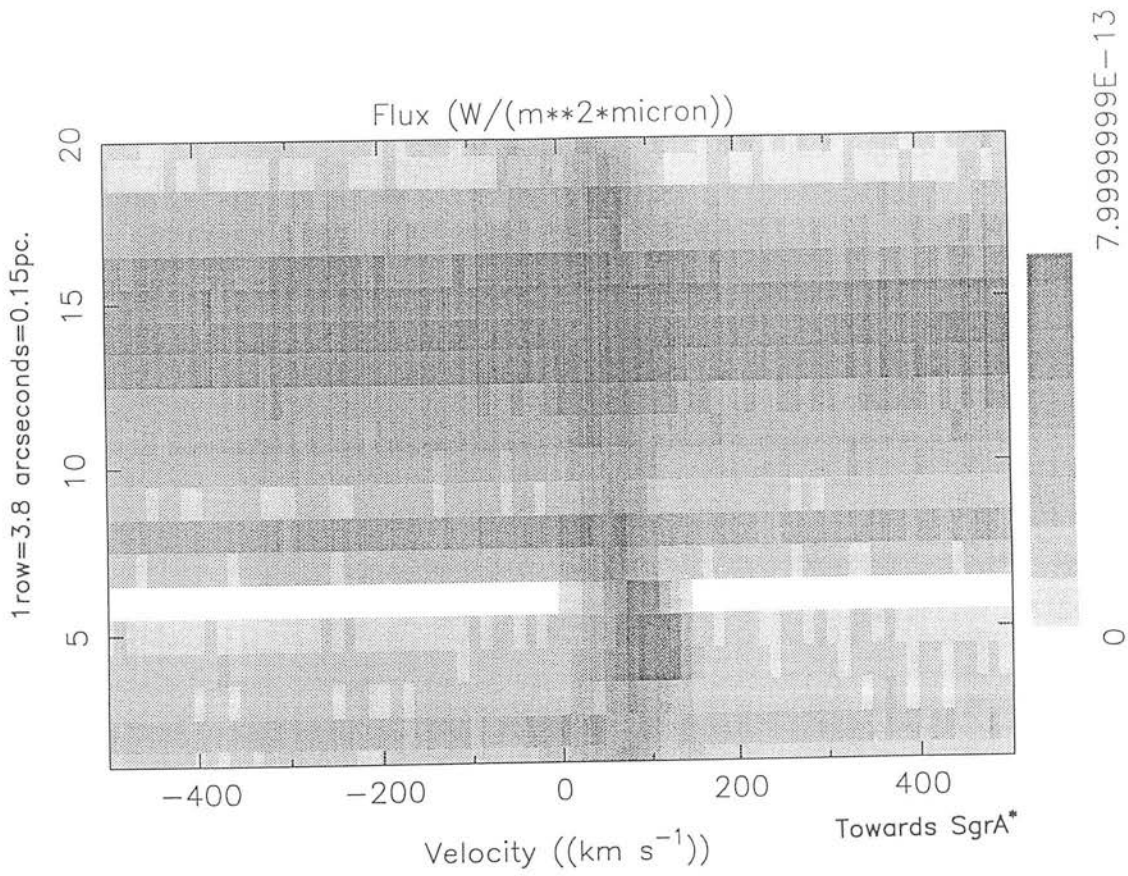


Figure 6.6: Position–velocity diagram of the northeast lobe of H_2 emission at the galactic centre.

4.1$\widehat{''}$S–1.5$\widehat{''}$N. This is coincident with the smaller enhancement seen to the SW of SgrA* in the low dispersion spectra. In the echelle spectra it appears to be more extended ($\sim 10\widehat{''}$) than in the low dispersion spectra ($\sim 3\widehat{''}$); this may be is due to the difficulties of observing the H_2 lines against a strong continuum with low resolution spectra. The area of strong emission seen to the NE of SgrA* in the low dispersion data is present in the echelle data, but is not the brightest area of emission. The brightest patch is centred around 17.0$\widehat{''}$E, 27.9$\widehat{''}$N. The corresponding spectra from the low–dispersion data are heavily contaminated by the Ca, Na and CO absorption lines alluded to in Section 6.2.1. In all cases the echelle spectra and low–resolution spectra are compatible.

In common with other data sets, we observe clumps of gas whose velocities imply that they are not associated with the overall motion of the ring. In Figure 6.7 a streamer of gas which is distinct from the motion of the molecular ring and has a velocity gradient

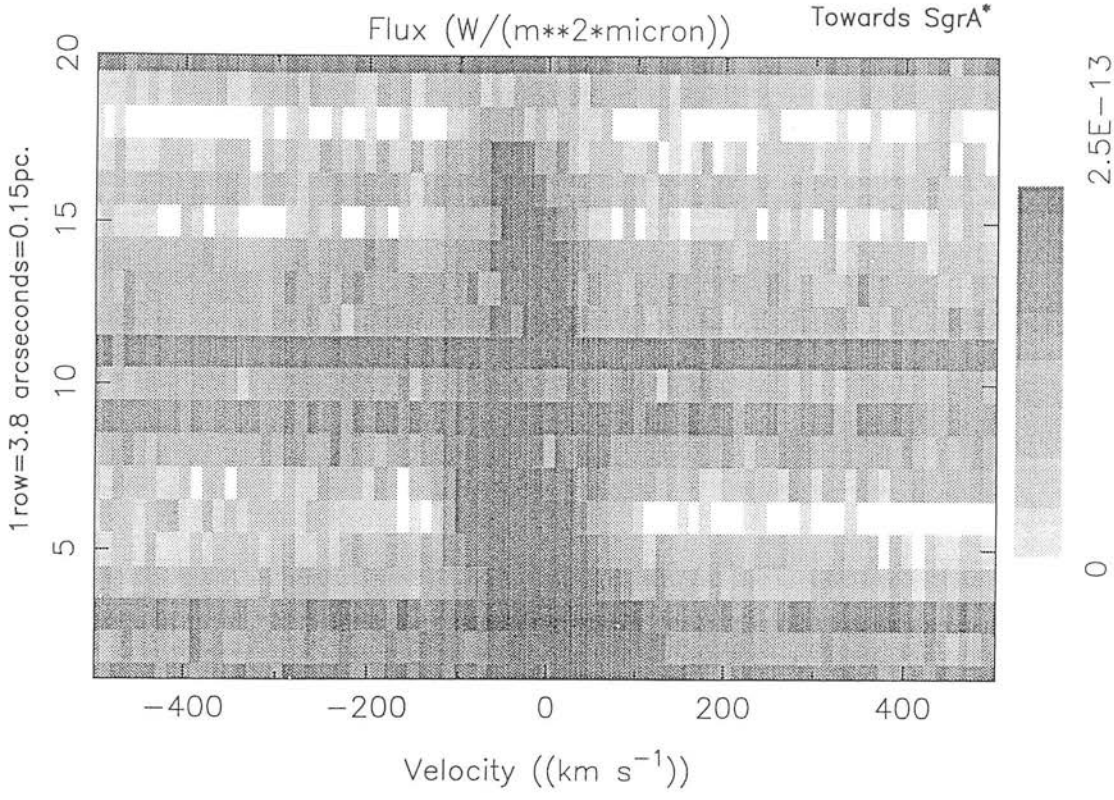


Figure 6.7: Position–velocity diagram of the southwest lobe of H_2 emission at the galactic centre.

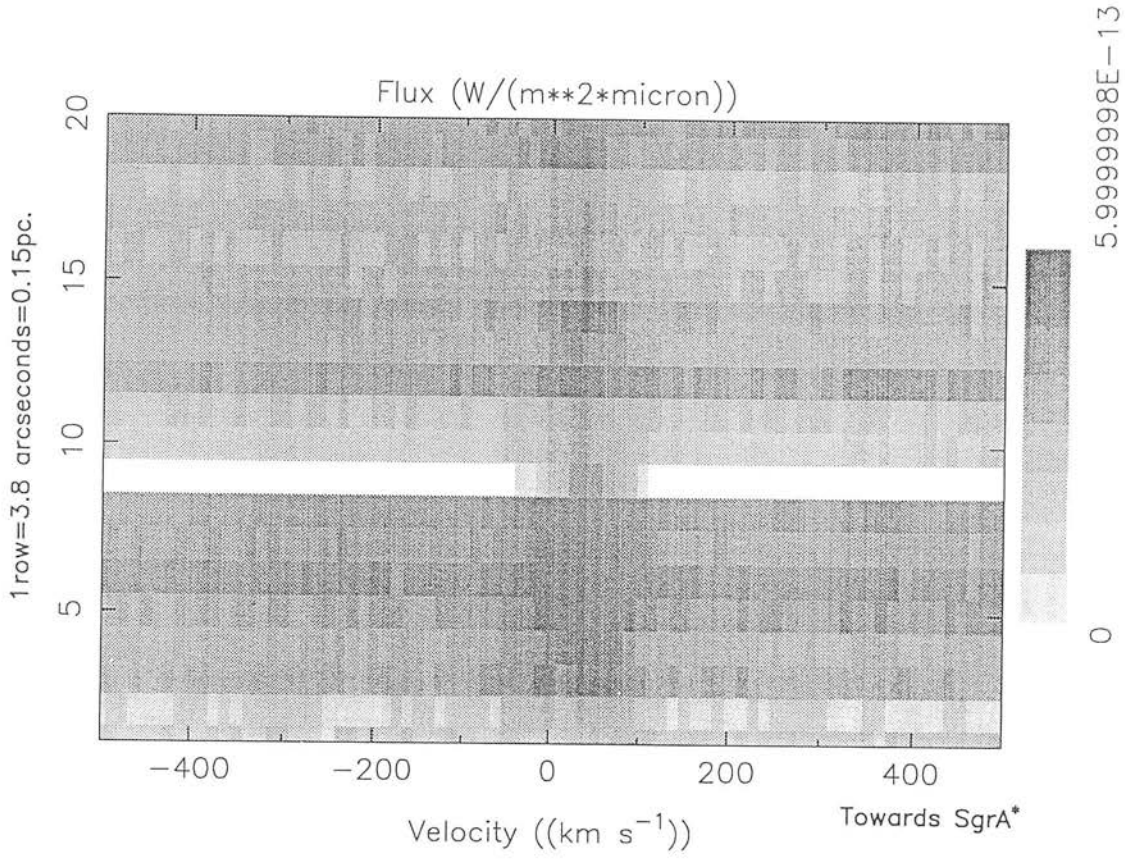


Figure 6.8: Position–velocity diagram of the northwest lobe of H_2 emission at the galactic centre.

greater than that of the ring can be seen extending from 3.9$\hat{'}$W, 4.3$\hat{'}$N to 9.7$\hat{'}$W, 13.3 $\hat{'}$S of SgrA*. Although this streamer is spatially coincident with the infrared source, IRS6, which appears in both the 10.6μm map which traces the warm dust (Rieke *et al.* 1978, Figure 6.9(a)) and the 15GHz map which traces the ionised gas (Figure 6.9(b)), the dynamics of IRS6 are quite different (Lacy *et al.* 1980).

A further bright clump of emission is seen with an approximately constant red–shifted velocity of $\sim$50km s^{-1} , extending over 10$\hat{'}$ at a radius of 2pc from the centre and lying in an area between 45$\hat{'}$ and 35$\hat{'}$W, 40$\hat{'}$ to 30$\hat{'}$S. Anomalous features with this velocity have been identified by Jackson *et al.* (1990) in their HCN data and they speculate that the material may be streamers from the +50km s^{-1} giant molecular cloud (Genzel *et al.* 1990; Zylka, Mezger and Wink 1990).

6.3 Discussion

6.3.1 Line ratios and excitation mechanism

Gatley *et al.* (1984) made the original measurement of the H_2 line ratios from the vibrational–rotational emission in the near IR and deduced that this gas was being shock excited by the wind from a mass–loss source. Recent advances in the theory of the UV excitation of H_2 (Burton, Hollenbach & Tielens, 1989), outlined in Chapter 5 lead us to understand that the spectrum of dense, fluorescently excited gas may emulate the characteristics of thermally excited gas, calling into question this initial diagnosis. Burton & Allen (1993) fit their H_2 data, along with fine structure line ratios, with a model of a clumpy PDR in which 65% of the gas has density 10^5cm^3, 13% is in clumps of density 10^7cm^3 and the remainder is at low density, taking a UV field strength of $10^5 G_o$. The hypothesis that the H_2 may be radiatively excited by the strong UV field from the centre is further supported by the close association of the H_2 gas with other species excited in star–forming regions *e.g.* [CII] (Lugten *et al.* 1987), [OI] (Jackson *et al.* 1992) and CO (Harris *et al.* 1985, Serabyn *et al.* 1986 and Sutton *et al.* 1990).

The measured ratio of the 1–0 S(1) and the 2–1 S(1) lines is 9.23$\pm$0.86; the rota-

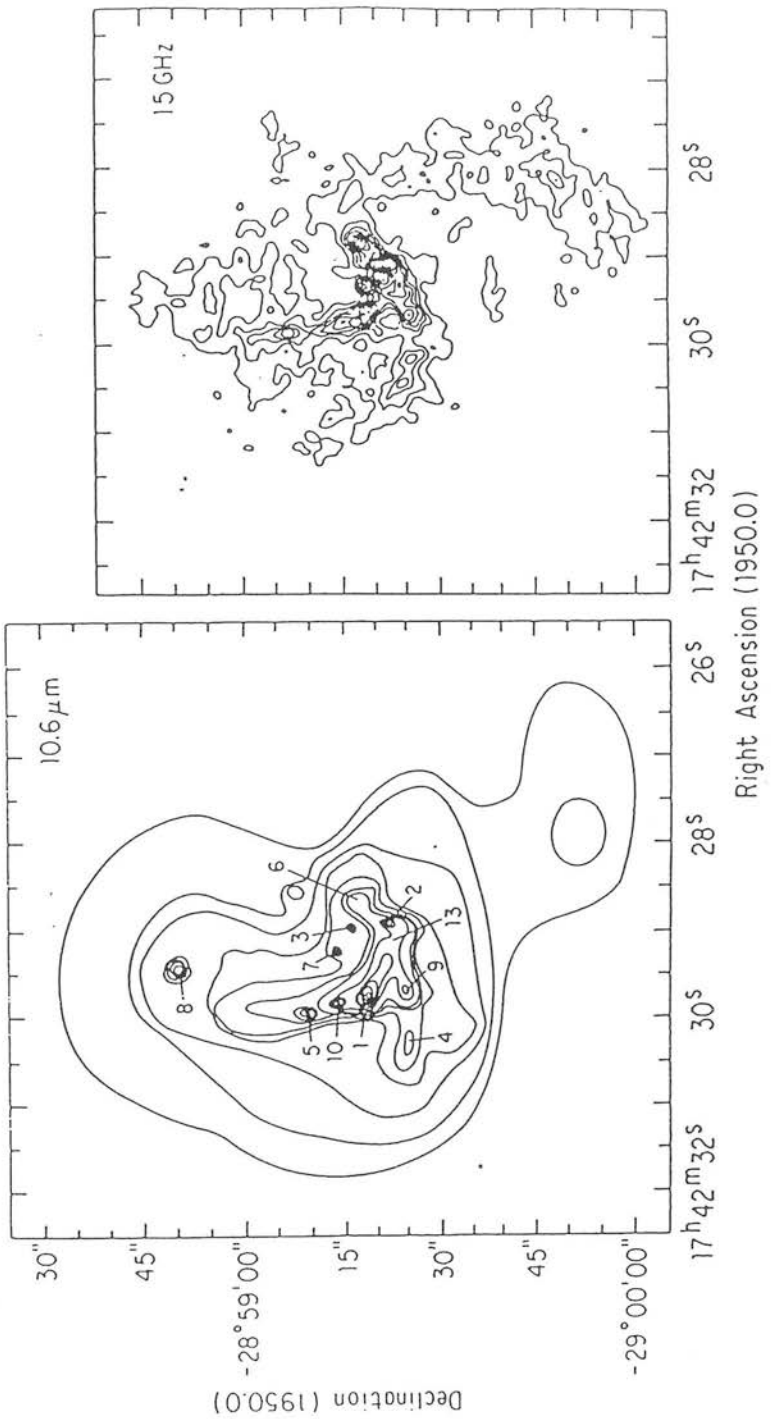


Figure 6.9: Two maps of the central few parsecs of the galaxy (taken from Brown and Liszt 1984). (a) shows the 10.6μm continuum emission from warm dust and the location of the infrared sources IRS 1–13 (Rieke *et al.* (1978)), (b) shows the 15Ghz emission which traces the ionised gas, showing clearly the structure of the streamers of ionised gas in the cavity.

tional and vibrational excitation temperatures are 1404K and 2220K respectively. The measured ortho/para ratio is 3.04. Comparing this line ratio with the models of Burton, Hollenbach & Tielens (1990) we find that to attribute this ratio to fluorescent excitation requires gas densities in the range $5\times10^5 cm^{-3}$–$1\times10^7 cm^{-3}$ coupled with UV field strengths of $\sim 3 \times 10^4 - 10^5 G_o$. The line intensity for the 1–0 S(1) line for such a model is predicted to be $\sim 2 \times 10^{-4}$ ergs s^{-1} cm^{-2}, in good agreement with the measured values which lies in the range 6.6×10^{-5}–5.3×10^{-4} ergs s^{-1} cm^{-2}. Evidence from other observations suggests that the required density and UV field conditions do exist in the molecular ring: Davidson *et al.* (1992) estimate the luminosity at the centre to be $10^7 L_\odot$. Assuming that half of these photons are capable of exciting the H_2 (Hollenbach 1990), this gives $I_{uv}= 5\times10^4 G_o$. In addition, densities for H_2 of $\sim10^5 cm^{-3}$ are implied from the data of Harris *et al.* (1985) and Gusten *et al.* (1987). Moreover, strong evidence for a clumpy structure to the ring implies the much higher density concentrations will be present.

In Shull and Hollenbach's (1978) model for H_2 which is excited by a shock in which the shock–speeds are not large enough for the gas to be dissociated, the line ratios for gas of density $3\times10^5 cm^{-3}$ and v_s=10km s^{-1} give the best agreement with our data for the relative intensities of the 1–0 S(1) line to the 1–0 S(0), 1–0 S(2) and 2–1 S(1) lines (Table 6.4). Again, this gas density is within the limits provided by other observations. The predicted absolute intensity for the 1–0 S(1) line is 5.46×10^{-3} ergs s^{-1} cm^{-2}, considerably brighter than the observed de–reddened value, which is in the range 6.6×10^{-5} – 5.3×10^{-4} ergs s^{-1} cm^{-2}. Thus, the possibility that the emission arises from gas which has been excited by collisions in a dense shock is not excluded.

The non–detection of the H_2 lines in the H–window does not contradict the possibility that the emission may be radiatively excited, although a feature of fluorescent emission vs. thermal emission is the enhancement of the emission lines from higher vibrational levels, *e.g.* those seen in the H–band relative to a thermal distribution (Figure 1.8). For the density and UV field implied for the galactic centre transitions from vibrational levels above v=1 will not be thermalised (Figure 5.6), thus the ratios of the higher v lines will show the fluorescent ratios from Black & van Dishoeck's Model 14. The strength of the 6–4 Q(1) line can be estimated relative to the 2–1 S(1) line since their ratio should be

Line	Flux relative to 1–0 S(1)	
	measured	predicted
1–0 S(0)	0.23±0.04	0.22
1–0 S(2)	0.32±0.04	0.36
2–1 S(1)	0.11±0.01	0.10

Table 6.3: Measured strengths of the 1–0 S(0), 1–0 S(2) and 2–1 S(1) lines compared with the predicted values for a model of gas of density $3\times10^5 cm^{-3}$ excited by a shock with v_s=10km s^{-1} (Shull & Hollenbach 1978).

1.7. The extinction to this line at 1.601μm will be ~5 magnitudes, significantly greater than that to the 2–1 S(1) line at 2.248μm , given our adopted extinction of A_v=30mag, so the observed line ratio would be 15. Therefore, the flux of the 6–4 Q(1) line, for a fluorescent model, would be $\sim6\times10^{-20}$ Wm^{-2}. From the sensitivities of CGS4, a 3σ detection of a line of flux $5.5\times10^{-19}Wm^{-2}$ requires 30min; thus a 3σ detection of the 6–4 Q(1) line requires ~4hrs. The spectra discussed were the result of twenty minutes integration on–source, so the non–detection presents no conflict to a fluorescent model.

To summarise these findings: the line fluxes, the 1–0 S(1)/2–1 S(1) line ratio, ortho/para ratio and excitation temperatures, and the non–detection of emission from the H band spectra, are all consistent with either fluorescent emission from a PDR of density in the range $5\times10^5 cm^{-3}$–$10^7 cm^{-3}$ illuminated by a UV field of strength in the range $10^{4-5}G_o$ or by emission from dense ($3\times10^5 cm^{-3}$) H_2 excited by a 10km s^{-1} , non–dissociative shock.

6.3.2 Molecular gas within the cavity

Both the low–dispersion and high–dispersion spectra show emission from H_2 arising from gas which appears at a projected radius smaller than the edge of the ring (1.54pc), suggesting that molecular material may exist within the central ionised cavity surrounding the neutral ring. In this Section, the possibility of the molecular gas existing in the

cavity is explored.

It has been proposed (Lo & Claussen 1983, Serabyn *et al.* 1988) that the presence of such gas could be the consequence of an in–fall scenario in which material from the molecular ring is accreted onto the central source and is being ionised as it approaches. Further evidence of this behaviour has been inferred from the observations of Harris *et al.* (1985) and Jackson *et al.* (1992). Harris *et al.* observed the CO 7–6 emission line with a width of 200km s^{-1} near the centre, which they took to be from neutral clumps which are being ionised and which are dynamically separate from the disk. From observations of the [OI] 63μm line Jackson *et al.* deduced that the amount of atomic gas within the disk was as much as twenty times the amount of ionised gas. Their map of the cavity shows a ridge of [OI] emission which is bordered by the eastern and northern arms; it is suggested that the arms are the ionised edge of a neutral cloud which has been accreted from the disk. If matter is indeed accreted in this way, then molecular gas should also be present. However, the strong UV field may dissociate these molecules before they approach the central source. For the specific case of H_2 , there are two mechanisms which could preserve the molecule in this environment: the absorption of the UV photons by dust grains (dust–shielding) or by other hydrogen molecules (self–shielding). Early studies of the dust at the galactic centre, by mapping the FIR continuum radiation implied that dust was present only in the ring. However, recent polarisation studies (Aitken *et al.* 1991) have shown that dust is present in the cavity, and that it is well mixed with the ionised gas. Marr *et al.* (1992) have shown that this dust can provide sufficient shielding to preserve the H^+CO molecule which they observe in absorption as close as 0.2pc from the ionising source.

If we hypothesise that H_2 gas is present within 0.4pc of the centre as observed, then the nature of that gas can be considered as follows. The galactic centre may be thought of simplistically as a UV source surrounded by a molecular cloud. Through the action of the UV photons emitted from that source, an HII region is formed, the size of which is determined by the electron density of the gas (n_e), the flux of ionising photons from the UV source (L_*) and the recombination coefficient for H_2 ($\alpha(T)$). The HII region will stop encroaching upon the molecular cloud when the recombination rate is balanced by the ionisation rate; at this point the radial distance from the source is the Strömgren

radius, (R_s), calculated from:

$$R_s n_e^{\frac{2}{3}} = \left[\frac{3L_*}{4\pi\alpha(T)}\right]^{\frac{1}{3}} \tag{6.1}$$

The flux of photons which are capable of ionising the H_2 (ie $E_{ph} \geq 13.6$ eV) has been determined by Geballe *et al.* (1984) to be 4×10^{49} per second. The electron densities inside the cavity have been estimated by Lo & Claussen (1983) to be $\sim 10^3$ for the low–density gas and $\sim 5\times 10^4$ for the denser clumps. Using the low density estimate implies a Strömgren radius of 0.96pc, which is rather smaller than the generally held value for the radius of the molecular ring (1.5pc), but not inconsistent with it.

For a clump of H_2 falling into the cavity, the UV flux at the radial distance of the clump from the centre can be estimated and used to determine a lower limit for the density of the clump. From the work of Burton, Tielens & Hollenbach (1990) it is found that the H_2 gas in a PDR will exist in one of two conditions: either it is in the regime where dust shielding dominates the emission and structure of the PDR, or the density is high enough that the gas is self–shielding. These two cases have been discussed more fully in Chapter 5, but to recap, in the case where

$$\frac{n}{G_o} \sim 1 \tag{6.2}$$

the H_2 is shielded by the absorption of the UV photons by dust, whereas if

$$\frac{n}{G_o} \geq 40 \tag{6.3}$$

the H_2 is self–shielding. A lower limit can be placed on the density by assuming that dust–shielding dominates. The total UV flux at the centre is estimated to be $\sim 10^7$ $L_\odot$ (Davidson *et al.* 1992), so that at a distance of 0.4pc, the projected distance of the clump of emission in the SW (Figure 6.7), the UV flux will be $6\times 10^5 G_o$. The implied lower limit to the density is thus 6×10^5 cm^{-3} (Equation 6.2). Burton & Allen (1993) modelled their H_2 line ratios from the ring using a two component model for the gas densities implying the presence of condensations with density up to $\sim 10^7$cm^{-3}. If streamers were to detach

from the ring, the low density gas would be quickly ionised leaving only the high density clumps. The peak intensity of observed emission from this clump is $\sim 6.22 \pm 1.8 \times 10^{-4}$ ergs s^{-1} cm^{-2} sr^{-1}, which is in excellent agreement with the value of 6.2×10^{-4} ergs s^{-1} cm^{-2} sr^{-1} which can be provided by a gas of density $10^6 cm^{-2}$ in a UV field of $10^5 G_o$ (Burton, Hollenbach & Tielens 1990).

6.3.3 Kinematics

Observations of the kinematics of the molecular ring, in particular the radial dependence of the gas velocity, have been used to study the underlying mass distribution. Support for a point like mass concentrated at the centre, possibly a black hole, has come from the CO J=7–6 observations of Harris *et al.* (1985) and the NeII observations of Serabyn and Lacy (1985), who agree on a mass of $\sim 4 \times 10^6$ $M_\odot$ for the central mass concentration from data which demonstrate a $r^{-0.5}$ velocity dependence. A number of workers observe no fall–off in velocity with increasing radial distance from the centre, among them Sutton *et al.*, who interpret their maps of CO J=3–2 as showing a rotational velocity of 110km s^{-1} over radii of 2pc to 8pc. This is in perfect agreement with the rotational velocity of 110km s^{-1} from 2pc to 5pc measured for HCN J=1–0 (Genzel *et al.* 1987) and HCN J=3–2 (Jackson *et al.* 1992). Serabyn *et al.* (1986) also observed no change in the rotational velocity in their observations of CO J=1–0. A hybrid model, consisting of a central point–like mass of 4×10^6 $M_\odot$ and a surrounding stellar distribution with mass 2×10^6 $M_\odot$ within the radius of the ring (1.7pc) is proposed by Lugten *et al.* to explain the observed velocity distribution of the [CII] 158μm line, giving $v_{rot} \propto r^{-(0.2-0.4)}$. The observations reported here are insufficient to specify fully a model of the H_2 emission as they only sample two slices across the area, but the data can usefully constrain with existing models.

Assuming that the data do indeed represent emission from a rotating, inclined ring, a correction can be applied to transform the data from velocities observed in our frame of reference to velocities observed in the galactic centre reference frame. To do this, a model of the velocities that would be observed from such a ring must be constructed. The geometrical relationships between the frames of reference are shown in Figure 6.10.

The treatment follows that from Mihalas & Binney's "Galactic Astronomy" (1968). The radial velocity measured at ρ,ϕ is

$$v_r(\rho,\phi) = v_o + \Pi(R,\theta)\sin\theta\sin i + \Theta(R,\theta)\cos\theta\sin i + Z(R,\theta)\cos i \tag{6.4}$$

where Θ, Π and Z are the radial and tangential velocities and the velocity perpendicular to the galactic plane and v_o is the velocity of the centre of mass. For the particular case of the galactic centre, the tangential velocity is known to be negligible (Jackson *et al.* 1992) as is the velocity out of the plane; the radial velocity is independent of position angle as the ring is axially symmetric. Correcting the velocities for the velocity relative to the Local Standard of Rest will reduce v_o to zero, greatly simplifying the Equation 6.4 to

$$v_r(\rho,\phi) = \Theta(R)\sin i\cos\left[\arctan\left(\frac{\tan\theta}{\cos i}\right)\right] \tag{6.5}$$

since

$$\tan\phi = \cos i\tan\theta \tag{6.6}$$

and where

$$R^2 = \rho^2\left(\cos^2\phi + \frac{\sin^2\phi}{\cos^2 i}\right). \tag{6.7}$$

The position angle, ϕ at which the observation was made is calculated from the RA and declination of the measurement, as is the radius from the dynamical centre which was defined as the distance from the position of SgrA* ($17^h\ 42^m\ 29.3^s$, -28° 59′ 18.5″ from Tollestrup *et al.* 1989). The form of $\Theta(R)$ is as yet undetermined. The data were corrected for the position angle of the observations, *i.e.* by the value of the angular terms of Equation 6.5; the resulting distribution of velocities with radius is shown in Figure 6.11 and the data are given in Table 6.3. Four models of the motion of the molecular ring are plotted on Figure 6.11, the parameters for which are given in Table 6.6 - the values for R_o have been scaled for a galactic centre distance of 8kpc. Of these previous models,

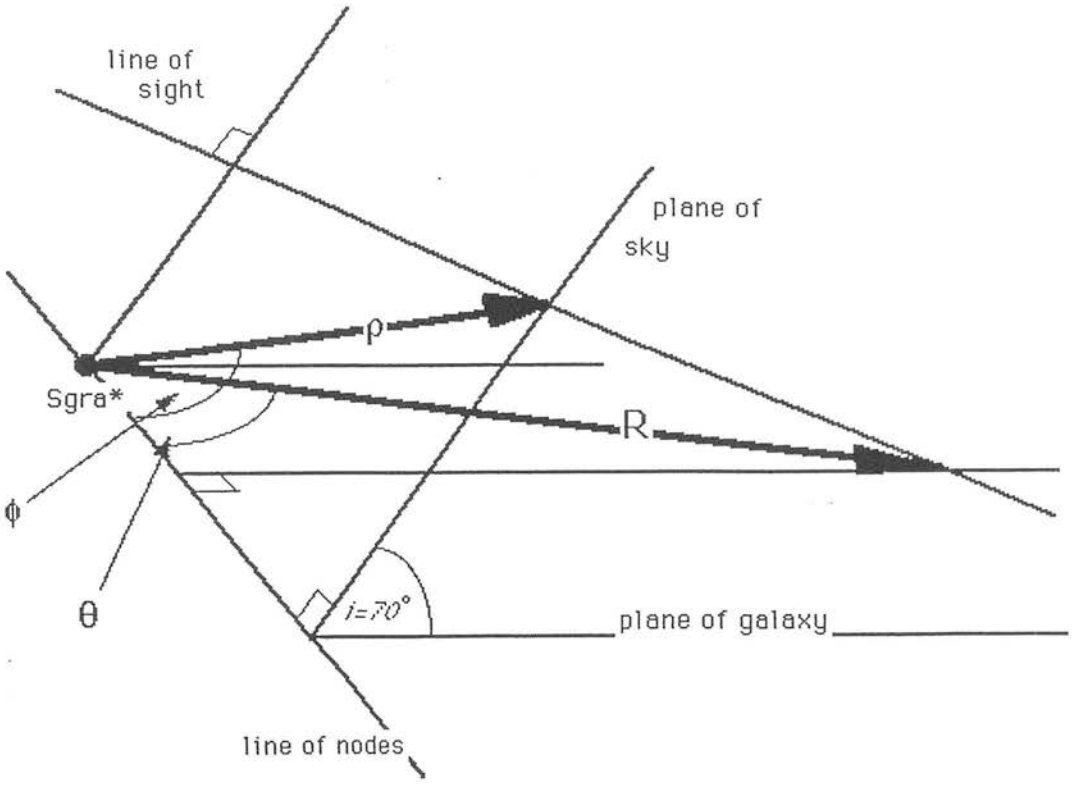


Figure 6.10: The relationship between the galactic plane and the plane of observations (after Mihalas & Binney 1968)

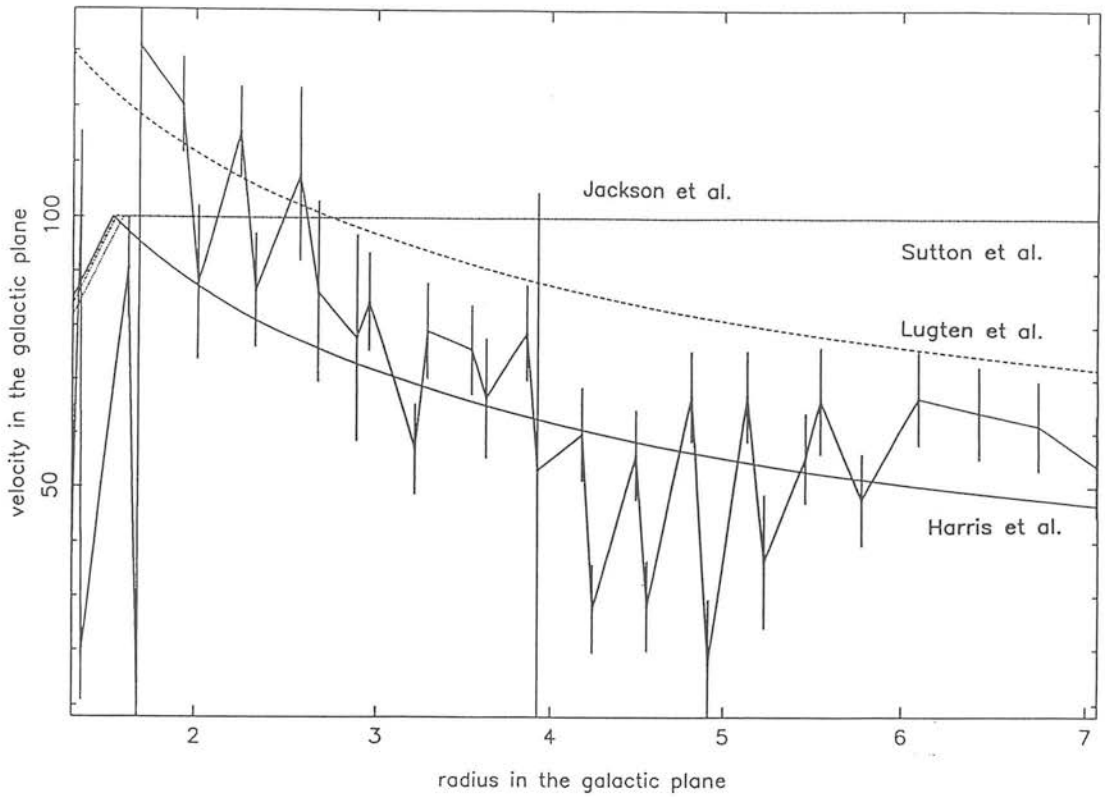


Figure 6.11: Velocity measured from the echelle spectra of the galactic centre as a function of radial distance from SgrA*. Superimposed on the data are the models of Harris *et al.*, Lugten *et al.*, Sutton *et al.* and Jackson *et al.*.

the best agreement with the data is obtained with a Keplerian velocity distribution, as measured by the Harris *et al.* (1985) observations of the molecular ring and the Lacy *et al.* fit to the ionised gas in the western arc.

Observations of the line widths of the H_2 lines can provide a means of differentiating between the shock and radiative excitation mechanisms (see for example Burton *et al.* 1990). Radiatively excited gas will have line widths equal to the thermal velocity dispersion of the cloud ($\sim$3km s^{-1}). In contrast, broad emission lines are observed from gas excited by a shock. To determine to what extent the observed H_2 widths at the galactic centre are due to the rotation of the ring, the following analysis was performed.

ρ (pc)	ϕ (° E of N)	$v_r(\rho,\phi)$ (km s^{-1})	R (pc)	$\Theta(R)$ (km s^{-1})
0.22	37.8	40.7 ± 4.1	0.43	55.42± 9.42
0.25	73.7	39.4 ± 1.8	0.70	77.20± 8.59
0.27	-77.2	-37.24 ± 1.9	0.77	78.98 ±9.26
0.28	5.38	35.7 ± 3.0	0.29	39.04 ±8.50
0.35	95.1	32.6 ± 2.0	1.02	82.82 ±9.13
0.39	-12.3	34.3 ± 2.7	0.45	72.80± 9.43
0.48	106.2	35.6 ± 2.7	1.35	104.91 ±10.61
0.49	-109.2	-16.46 ±1.7	1.36	19.36 ± 8.50
0.52	-21.7	43.8 ± 2.3	0.74	193.93±12.15
0.62	112.6	40.1 ± 4.1	1.69	130.85± 14.34
0.63	-115.6	-10.24 ±1.5	1.68	8.32± 8.35
0.71	21.6	66.5 ±3.0	1.01	72.56 ±8.56
0.76	116.6	22.8 ± 3.7	2.02	87.94±14.18
0.85	25.3	31.1 ± 3.1	1.31	38.11 ±8.64
0.90	119.3	20.6 ± 2.0	2.34	86.56 ± 10.48
1.00	27.9	78.4 ± 4.0	1.63	90.62 ± 9.10
1.05	121.3	19.17 ± 4.1	2.68	86.33 ±16.69
1.14	29.8	103.4 ± 2.7	1.93	120.32± 8.54
1.23	-127.2	-64.67 ± 3.0	2.96	84.53 ± 9.06
1.29	31.4	97.2 ± 1.1	2.25	115.48 ± 8.10
1.38	-128.2	-61.42±2.8	3.28	79.13 ± 8.91
1.44	32.7	89.02 ± 11.7	2.58	107.79 ±15.62
1.54	-129.1	-52.94± 5.6	3.63	66.64 ±11.16
1.59	33.6	62.2 ± 15	2.89	77.82 ±19.12
1.68	-129.8	-43.60 ± 36.8	3.92	53.28 ± 51.42
1.74	34.4	43.91 ± 2.0	3.21	57.14 ± 8.33
1.83	-130.4	-25.22± 1.0	4.24	27.74 ± 8.12
1.89	35.2	59.18 ± 2.1	3.54	75.65 ± 8.37
1.98	-130.9	-25.65 ± 1.7	4.56	28.19 ± 8.33
2.04	35.8	61.45 ± 3.4	3.86	78.84 ± 8.96
2.14	-131.2	-17.87 ± 6.5	4.91	17.51 ± 11.97
2.19	36.3	45.19 ± 2.7	4.18	59.88 ± 8.62
2.29	-131.6	-31.87 ± 6.9	5.23	36.42 ±12.31
2.34	36.7	41.79 ± 1.7	4.50	56.07 ± 8.26
2.44	-132.0	44.06 ± 4.2	5.55	66.23 ± 9.80
2.49	37.1	50.56 ± 2.0	4.82	66.87 ± 8.35
2.64	37.4	50.42 ± 2.1	5.14	66.92 ± 8.39
2.79	37.8	40.80 ± 1.7	5.46	55.54 ± 8.26
2.94	38.0	34.4 ± 2.3	5.78	47.88 ± 8.47
3.09	38.2	49.7 ± 2.7	6.09	66.62 ± 8.65
3.24	38.5	47.3 ± 2.5	6.42	63.91 ± 8.56
3.39	38.7	45.2 ± 1.8	6.74	61.47 ± 8.30
3.54	38.9	39.0 ± 4.0	7.06	54.00 ± 9.38

Table 6.4: The measured velocities for the H_2 at the Galactic centre. The velocities corrected to the plane of the galaxy are given, along with the de–projected radii.

Galactic Centre Velocity Models			
ring edge R_o (pc)	velocity at ring edge Θ_o (km s^{-1})	Radial velocity dependence α	reference
1.54	100	-0.50	Harris *et al.* 1985
1.30	130	-0.35	Lugten *et al.* 1986
1.60	110	-	Sutton *et al.* 1990
1.56	110	-	Jackson *et al.* 1992

Table 6.5: The parameters for the four models for the dynamics of the Galactic centre plotted. The R_o are scaled to a galactic centre distance of 8kpc.

Adopting the radial velocity function

$$\Theta(R) = \Theta_o \left[\frac{R_o}{R}\right]^{0.5} \tag{6.8}$$

for a ring with its rotation axis inclined at 70° to the line of sight and with a position angle on the sky of 27° (*e.g.* Jackson *et al.* 1992) with Θ_o=100km s^{-1} and R_o=1.54pc, the expected dispersion at a given ρ,ϕ can be calculated and used to compare the line widths due to the rotation of the ring with those observed. The dispersions are separated into two orthogonal components: the velocity dispersion as a function of radius and that as a function of position angle, which are calculated from the derivatives of the velocity distribution (Equation 6.5)

$$\frac{\delta v_r(\rho,\phi)}{\delta\rho} = -0.5\frac{vel_{mod}}{\rho} \tag{6.9}$$

$$\frac{\delta v_r(\rho,\phi)}{\delta\phi} = 0.5 vel_{mod}\frac{\sin^2\phi\sin^2 i}{\left(\cos^2\phi\cos^2 i + \sin^2\phi\right)^2} \tag{6.10}$$

where vel_{mod} is the velocity in the plane of the sky for this model. These dispersions will give the actual width when multiplied by the angular/radial extent of a given pixel, found from the geometry of the observations. This process has been complicated by the fact that the slits were placed across the peaks of H_2 emission, estimated from the Gatley *et al.* (1986) map (Figure 6.1), rather than symmetrically across SgrA*, in which case only two position angles would have been observed. To find the total line profile, these two components and that due to the instrumental resolution, all of which have triangular profiles, must be convolved. Numerically it was found that the convolution of two triangles has a FWHM equal to the quadratic sum of the FWHM of the individual triangles. The total value is given in Table 6.5, along with the observed line width. The model and observed dispersions are also shown in Figure 6.12. The Figure clearly shows that the observed line widths do not display any trend with decreasing distance from the centre, as is demonstrated by the model, but appears to remain approximately constant. For positions close to the ring, the model dispersion is far greater than that observed. For radial distances of greater than ~3pc, the gas is strongly associated with the molecular ring and the theoretical dispersion is in good agreement with the observed

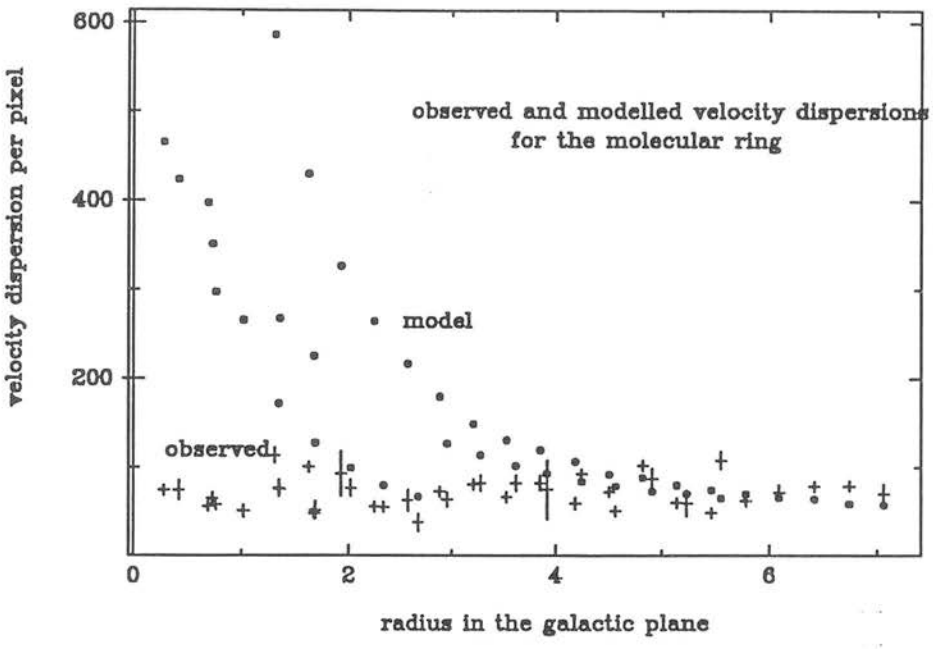


Figure 6.12: Model and observed dispersion for the H_2 emission lines at the galactic centre, using the Harris *et al.* model for the ring dynamics.

line widths. The most credible explanation for this discrepancy between the model and the observations for small projected radii is that the gas which is observed at a projected radius within the central cavity is in fact foreground gas. The fact that the velocity dispersion observed is the same as that of the molecular gas from the ring implies that this gas is orbiting the galactic centre at the same radius as the ring.

The observed line width can thus be explained by the velocity dispersion due to the motion of the gas, from which it can be concluded that shock excitation plays no part in or makes a negligible contribution to, the excitation of the gas within the ring.

ρ (pc)	ΔV_{model} (km s^{-1})	$\Delta v_r(\rho\phi)$ (km s^{-1})	ρ (pc)	ΔV_{model} (km s^{-1})	$\Delta v_r(\rho\phi)$ (km s^{-1})
0.22	423.28	74.2 ±10.6	1.54	101.50	81.9 ± 8.7
0.25	397.20	56.0 ± 4.7	1.59	179.48	72.5 ± 3.3
0.27	297.00	57.6 ± 4.3	1.68	92.74	74.6 ±32.9
0.28	465.18	73.9 ± 6.3	1.74	148.64	80.5 ± 5.6
0.35	265.46	50.6 ± 6.6	1.83	84.27	92.5 ± 1.9
0.39	894.50	68.3 ± 8.7	1.89	130.46	66.6 ± 5.6
0.48	171.95	76.5 ± 8.9	1.98	78.49	50.6 ± 5.4
0.49	267.57	75.3 ± 4.7	2.04	119.36	82.2 ± 8.2
0.52	351.07	65.0 ± 6.4	2.14	72.70	87.1 ±11.3
0.62	127.24	51.6 ± 9.6	2.19	106.35	59.6 ± 6.4
0.63	225.22	48.0 ± 3.9	2.29	69.70	59.3 ±13.9
0.71	839.05	100.7 ± 7.3	2.34	91.85	72.3 ± 4.7
0.76	99.12	76.5 ± 8.9	2.44	65.06	107.8 ± 9.6
0.85	586.67	113.4 ± 8.2	2.49	88.33	101.9 ± 5.4
0.90	79.47	55.3 ± 4.2	2.64	79.50	59.8 ± 4.9
1.00	430.07	100.3 ± 7.8	2.79	74.12	48.5 ± 4.2
1.05	66.96	37.9 ± 9.6	2.94	69.68	62.1 ± 5.6
1.14	326.69	92.5 ±24.8	3.09	65.64	71.6 ± 8.0
1.23	126.66	63.3 ± 7.8	3.24	64.11	78.2 ± 5.2
1.29	264.50	55.6 ± 4.7	3.39	58.51	78.6 ± 4.7
1.38	113.72	81.9 ± 8.7	3.54	57.52	70.2 ±10.4
1.44	216.47	62.6 ±11.5			

Table 6.6: The model and observed velocity dispersions for the galactic centre H_2 emission. The ring edge was taken to be 1.54pc, at which the velocity is 100.0km s^{-1} (Harris *et al.* 1985); the ring is inclined at 70° with a position angle of 27°E of N (Jackson *et al.* 1992).

6.4 Conclusions

It has been shown that radiative excitation of a high density gas by UV photons from young stars at the galactic centre is sufficient to explain observations of the H_2 ro–vibrational emission from the molecular ring surrounding the galactic centre. The two observations which must be explained by this model are the 1–0 S(1)/2–1 S(1) line ratio measured for the H_2 emission from the molecular ring and the broad line widths, both of which are usually associated with shock–excited gas,

H_2 line ratios determined from emission within the ring yield values for the rotational and vibrational excitation temperatures of 1404K and 2220K respectively; the 1–0 S(1)/2–1 S(1) ratio is 9.23± 0.86. The strength of the H_2 emission may be provided by fluorescent excitation of a gas of density $5\times10^5 cm^{-3}$–$10^7 cm^{-3}$ illuminated by a UV source of $10^{4-5} G_o$. However, these observations may also be explained by collisional excitation of a gas of density $3\times10^5 cm^{-3}$ due to the passage of a 10km s^{-1} shock. These density and UV field conditions have been inferred for the galactic centre from previous data sets.

The spatial and velocity distributions of the H_2 gas in the ring confirm Harris *et al.*'s model of the ring (1985), having radius 1.54pc and a velocity at the edge of 100km s^{-1}, decreasing as $R^{-0.5}$. A central, point–like mass of 4×10^6 $M_\odot$ is implied. Those models which show a uniform velocity as a function of distance from the centre are excluded by theses data. Analysis of the velocity dispersions of the gas shows that broad line widths observed can arise due to the distribution of velocities in a finite pixel. Thus the gas may be excited purely by the UV photons from the centre, as no thermal broadening is required to explain the line widths.

The possibility that emission from H_2 occurs from clumps of gas within the radius of the ring has been investigated. A lower limit to the density of this gas (6×10^5 cm^{-3}) is determined by the UV intensity at the projected radius of the emission ($6\times10^5 G_o$). The measured line intensity for the 1–0 S(1) line is in good agreement with the model intensity for a gas of this density illuminated by a UV field of this strength. However, the measured velocity dispersions of the gas from the data presented suggest that the

gas is foreground material orbiting SgrA* at the same radial distance as the ring.

Chapter 7

Conclusions

7.1 Summary

This work of this thesis has been concerned with the development of a new infrared array spectrometer, Cooled Grating Spectrometer 4, and its application to astronomy. The motivation behind the development of any astronomical instrument is the desire to apply new technologies to astronomical problems, and the design of the instrument is strongly influenced by the requirements of the observational programmes. The results from two astronomical projects carried out with CGS4, observations of the H_2 emission from the planetary nebula Hubble 12 and from the Galactic centre have been reported in Chapters 5 and 6. The question at the end of this study must be: has CGS4 worked?

7.2 Fulfillment of the Design Goals

The build a sensitive spectrometer requires both high throughput and good image quality; the success of CGS4 can be measured from the image quality and sensitivities achieved, and their relation to the predicted values.

7.2.1 Image Quality

In Chapter 2 the detailed optical design of CGS4 was presented. It was driven by the need for high throughput and excellent image quality. The design spot sizes are $\sim 30\mu$m , chosen to be optimised for the 76μm pixels of the 58×62 detector array. but also in anticipation of the imminent availability of the 30μm pixels in a 256×256 detector array. The degradation in image quality for the 76μm pixels should be less then 10%.

The tolerances on manufacture and alignment of the optical components to meet this goal, and the need to cool the instrument to LN_2 temperatures to minimise the dewar background, necessitated the use of an all aluminium construction for the optics, to eliminate differential contraction on cooling, diamond machining for the mirror manufacture. This approach lead to a flexible, modular design of the instrument, allowing, for example, the two cameras to be interchanged without re–alignment of the camera mirrors. Tests described in Section 3.6 show that the alignment of any module to the optical support structure is maintained if that module is removed and replaced, so that re–alignment of the camera module to the casting is also unnecessary. The techniques used originally to align the optics were detailed in Chapter 3. The proof of the effectiveness of this approach is seen in the image quality of the final, assembled instrument, and in the sensitivities measured for the spectrometer. The alignment was also shown to remain within the tolerances after the instrument had been cooled and re–warmed. The "before" and "after" measurements for the tolerances are given in Table 7.1.

The image quality measured at NIR wavelengths was reported in Chapter 3. The FWHM of unresolved lines is 1.02 pixels at the centre of the array, with variation of typically 0.01 pixels seen at the ends of the slit and the extremes of the array. Thus the 30μm spot sizes have been achieved for the NIR, both for laboratory spectra and for astronomical observations.

7.2.2 Sensitivity

A model to predict the sensitivity of a NIR instrument was outlined in Section 2.8.2, and the theoretical sensitivities found. Actual sensitivities for CGS4 at UKIRT were

observer as the efficiency for the non–optimum order is ~10%. Crucial to the smooth operation of this instrument was the need for the calibration to be repeatable over timescales of around 6months–1 year. Tests of the repeatability and reproducibility of the grating drives and translation mechanism showed that they were repeatable to 0.1pixel over the equivalent of one year of observations.

7.4 Flat–fielding a 2D Array Spectrometer

One of the problems identified for CGS4 was the need to provide a flat–field frame with which to calibrate the pixel–pixel gains for the 2D array. To this end, a tailormade calibration unit containing two continuum sources, a tungsten–halogen lamp and a 1173K black–body, as well as three emission line sources, was designed and constructed. The CU optics fill the slit plane with a field identical to that from the telescope secondary. A study of the resulting flat–fields from this source showed that a signal/noise ratio of 4×10^4 is achievable with a single frame, but that the accuracy with which any observation could be flat–fielded would be set by the temporal dependence of the pixel–pixel gains. A drift of 0.1% per night is expected. The flats were shown to be wavelength dependent, obliging the observer to obtain a flat–field for each new configuration.

An investigation of the success of the CU at providing a spatially flat–field, for removing the pixel–pixel variations along the slit, was carried out. The illumination of the array by the black–body was compared with the illumination from the sky, the flatness of the sky having been estimated at $:10^4$ for wavelengths in the 1–2.3μm region from theory of the structure of OH emission in the atmosphere (Chapter 4). It was seen that, at the 3σ level, there was no difference in the spatial profiles from these sources. Hence the design of the calibration unit and intergrating sphere have been successful.

7.5 Astronomical Results from CGS4

The NIR is an excellent waveband for observations of molecular species, and the two observational projects reported in this thesis have been observations of the rotational–

vibrational emission lines of H_2 from Hubble 12 (Chapter 5) and from the molecular ring at the galactic centre (Chapter 6).

7.5.1 Pure Fluorescent Emission from Hubble 12

A spectrum of the K–band (2.0–2.4μm) emission from Hubble 12 was observed with an on–source exposure time of 10mins. A total of 16 H_2 lines been detected in this source, including a number which had not previously, for example the 2–1 S(0), 3–2 S(1) and The column densities derived for these new, high–excitation lines, allow the unequivocal statement that the H_2 emission from Hubble 12 is pure fluorescent emission from a low–density gas. The H_2 density derived for the gas is 10^4–10^5cm^{-3}; it is illuminated by a UV field of $10^4 G_o$.

7.5.2 H_2 Emission from the Galactic Centre

As an extended, highly–obscured source the galactic centre is the ideal target for infrared long–slit spectroscopy. Previously, no data set had the advantages of both high spatial (0.1pc) and spectral (R$\sim$8000) resolution for observations of this region.

It has been shown the 1–0 S(1)/2–1 S(1) line ratio measured for the H_2 emission from the molecular ring (9.23$\pm$0.86) and the broad line widths ($\sim$50–100km s^{-1}), more usually associated with shock–excited gas, can be explained by radiative excitation of a high density gas by UV photons from young stars at the Galactic centre. The strength of the H_2 emission may be provided by fluorescent excitation of a gas of density 5×10^5cm^{-3}–10^7cm^{-3} illuminated by a UV source of $10^{4-5}G_o$. However, these observations may also be explained by collisional excitation of a gas of density 3×10^5cm^{-3} due to the passage of a 10km s^{-1} shock.

The spatial and velocity distributions of the H_2 gas confirms that the ring has radius 1.54pc and a velocity at the edge of 100km s^{-1} , decreasing as R$^{-0.5}$ (Harris *et al.* 1985). A central, point–like mass of 4×10^6 $M_\odot$ is implied. Those models which show a uniform velocity as a function of distance from the centre are excluded by these data. The velocity dispersions per pixel expected for observations of such a ring shows that

broad line widths observed can arise due to the distribution of velocities in a finite pixel. Thus the gas may be excited purely by the UV photons from the centre, as no thermal broadening is required to explain the line widths.

The data imply the presence of H_2 emission arising from clumps of gas within the radius of the ring. A lower limit to the density of this gas was determined (6×10^5 cm^{-3}) for a UV intensity at the projected radius of the emission of $6\times10^5 G_o$. The measured line intensity for the 1–0 S(1) line is in good agreement with the model intensity for this clump of gas. However, the measured velocity dispersions of the gas from the data presented suggest that the gas is foreground material orbiting SgrA* at the same radial distance as the ring.

7.6 Future Scientific Projects

The work on Hubble 12 has clearly demonstrated that it can be considered as a standard source against which all other fluorescent sources must be tested. There is much room for further investigation of this object to confirm the existing models of fluorescent emission, *e.g.* higher excitation lines in the H and J bands have yet to be observed. Observations of emission lines arising from transitions from levels as high as v=6 may expand our knowledge of the H_2 formation and its effect on the fluorescent emission spectrum. The reason for the ortho/para ratio being less than 3 for some fluorescent sources is unknown, but it may be set by formation conditions.

The Galactic centre is a constant source of questions. The work presented in Chapter 6 sets some strong constraints for the nature of the emission seen from the molecular ring, however it is still not possible to say what fraction of the gas owes its excitation to a shock and what fraction to radiative excitation. One of the abiding problems with observations of this region is the determination of the extinction. Observations of higher v levels are required before the fluorescence or shock–model can be effective distinguished. The calculation in Chapter 6 indicated that these observations are impossible with current technology, but with the 256^2 arrays, the required sensitivity may be achieved. The possible existence of foreground gas, implying a third dimension for the ring, prompts

further investigation to provide a full model of the dynamics of the H_2 . The confusion of dynamical models, streamers and clumps of gas within the molecular ring and the cavity are far from understood. It is unlikely that the nature of the central source will be revealed until they are.

7.7 Future developments in Instrumentation

In the two years which have passed since CGS4 was commissioned, the developments in instrumentation for the NIR have continued to gather pace. The 58×62 pixel array that provides CGS4 with it current sensitivity performance is soon to be superceded by the 256^2 InSb arrays that have become available. The first of these will be installed in the infrared camera, IRCAM, on UKIRT by late 1993. The design of CGS4 was generalised to include the possibility of replacing the 58×62 pixel array with the larger arrays as they became available, and the 256^2 array is due to be installed in the spectrometer in 1994. As well as the increased sensitivity due to a the multiplex advantage obtained by having 65536 pixels compared with 3596, the reduction in the pixel size by a factor of 6 gives a factor of 2.5 reduction in the background signal strength, hence a decrease in the limiting magnitude of 0.4, for background limited observations. The smaller pixels also have a smaller capacitance, which reduces the effective read–out noise for a single read from $400e^-$ with CGS4 to $50e^-$.

References

Aitken, D.K., Gezari, D., Smith, C.H., McCaughrean, M.C. & Roche, P.F. 1991, *Ap.J.*, **380**, 419.

Allen, D.A. 1989, in *The Center of the Galaxy*, IAU Symposium 136.

Allen, D.A., Hyland, A.R. & Hillier, D.J. 1990, *M.N.R.A.S.*, **244**, 706.

Backer, D.C. & Sramek, R.A. 1982, *Ap.J.*, **260**, 512.

Backer, D.C. & Sramek, R.A. 1987, in *The Galactic Center*, Proceedings of AIP Conference 155, 163.

Bates, D.R. & Nicolet, M. 1950, *J. Geophys. Res.*, **55**, 301.

Black, J.H. & van Dishoeck, E.F., 1987 *Ap.J.*, **322**, 412.

Brown, R.L. & Liszt, H.S. 1984, *Ann. Rev. Astrom. Astrophys.*, **22**, 223.

Burton, M.G. 1992, *Aust. J. Phys. submitted.*

Burton, M.G., Geballe, T.R., Brand, P.W.J.L. & Moorhouse, A. 1990, it Ap.J., **352**, 625.

Burton, M.G. & Allen, D.A. 1993, *in Astronomical Infrared Spectroscopy*, A.S.P. Conference Series, ed. S. Kwok. *in press.*

Burton, M.G., Hollenbach, D.J. and Tielens, A.G.G.M. 1990, *ApJ*, **365**, 620.

Chapman, A.R., Beard, S.M., Mountain, C.M., Pettie, D.G., Pickup, D.A. & Wade, R., 1990 in *Instrumentation in Astronomy VII*, ed. Crawford, D.L., (Proc.SPIE, 1235), 34.

Chamberlain, J.W., 1961. In *Physics of the Aurora and Airglow, International Geophysics Series*, Vol.2.. p.368.

Chrysostomou, A., Brand, P.W.J.L., Burton, M.G. & Moorhouse, A. 1992, *M.N.R.A.S.*, **256**, 528.

Davidson, J.A., Werner, M.W., Wu, X., Lester, D.F., Harvey, P.M., Joy, M. & Morris, M. 1992, *Ap.J.*, **387**, 189.

Dinerstein, H.L., Lester, D.F., Carr, J.S. & Harvey, P.M. 1988, *Ap.J.*, **327**, L27.

Draine, B.T. 1989, in Proceedings of the 22nd Eslab Symposium *Infrared Spectroscopy in Astronomy*, 93

Eckers, R.D., van Gorkom, J.H., Schwarz, U.J. & Goss, W.M. 1983, *A.A.*, **122**, 143.

Eckhart, A., Genzel, R., Krabbe, A., Hofmann, R., van der Werf, P.P. & Drapatz, S. 1992, *Nature*, **355**, 526.

Fowler, A.M., Probst, R.G., Britt, J.P., Joyce, R.R. & Gillett, F.C., 1987, *Optical Engineering*, **26**, 232.

Forrest, W.J., Shure, M.A., Pipher, J.L. & Woodward, C.E. 1987, in *The Galactic Center*, Proceedings of AIP Conference 155, 163.

Gatley, I., Jones, T.J., Hyland, A.R., Beattie, D.H. & Lee, T.J. 1984, *M.N.R.A.S.*, **210**, 565.

Gatley, I., Jones, T.J., Hyland, A.R., Wade, R., Geballe, T.R. & Krisciunas, K., 1986 *MNRAS*, **222**, 299.

Gattinger, R.L., 1969. *Ann. Geophys.*, **25**, 825.

Gattinger, R.L.,1971. In *The Radiating Atmosphere*, p.51, ed. M^{c}Cormack, B.M.

Geballe, T.R., Krisciunas, K., Lee, T.J., Gatley, I., Wade, R., Duncan, W.D., Garden, R. & Becklin, E.E., 1984 *Ap.J.*, **284**, 118.

Genzel, R., Watson, D.M., Crawford, M.K. & Townes, C.H. 1985., *Ap.J.*, **297**, 766.

Genzel, R. & Townes, C.H. 1987, *Ann. Rev. Astron. Astrophys.*, **25**, 377.

Genzel, R., Stacey, G.J., Harris, A.I., Townes, C.H., Geis, N., Graf, U.U., Poglitsch, A. & Stutzki, J. 1990, *Ap.J.* **356**, 160.

Gillingham, P.G. & Lankshear, A. 1990, in *Instrumentation in Astronomy VII*, ed. Crawford, D.L., (Proc.SPIE, 1235), 9.

Good, R.E., 1976, *Planet. Space Sci.*, **24**, 389.

Güsten, R., Genzel, R., Wright, M.C.H., Jaffe, D.T., Stutzki, J. & Harris, A. 1987, *Ap.J.*, **318**, 124.

Habing, H.J. 1968, *Bull. Astr. Inst. Netherlands*, **19**, 421.

Harris, A.I., Jaffe, D.T., Silber, M. & Genzel, R. 1985, *Ap.J.*, **294**, L93.

Hasegawa, T., Gatley, I., Garden, R.P., Brand, P.W.J.L., Ohishi M., Hayashi, M. & Kaifu, N. 1987, *Ap.J.*, **318**, L77.

Hatfield, R., Tuan, T.F. & Silvermann, S.M., 1981. *J. Geophys. Res.*, **86**, 2429.

Herschel, W. 1800, *Phil. Trans. R. Soc.*, **90**, 284.

Hersé, M., Thuillier, G., Camman, G., Chevassaut, J.–L. & Fehrenbach, M., 1989, *Appl. Optics*, **28**, 3944.

Heyvaerts, J., Norman, C. & Pudritz, R. 1988, *Ap.J.*, **330**, 718.

Hines, C.O., 1960. *Can. J. Phys.*,**38**, 1441.

Hines, C.O., 1974. *The Upper Atmosphere in Motion*, Geophysical Monograph No.18,

American Geophysical Union.

Hollenbach, D. 1990, in *The Evolution of the Interstellar Medium*, ed. L Blitz, A.S.P. Conference Series 12, 167.

Horne, K., 1986, *PASP* **98**, 609.

Howell, C.D., Michelangell, D.V., Allen, M. & Yung, Y.L., 1990. *Planet. Space Sci.*, **38**, 529.

Jackson, J.M., Geis, N., Genzel, R., Harris, A.I., Madden, S.C., Poglitsch, A., Stacey, G.J. & Townes, C.H. 1992, *Max Planck preprint.*

Jefferts, K.B., Penzias, A.A. & Wilson, R.W. 1973, *Ap.J.*, **179**, L57.

Kingslake, R. 1965, Ilumination in Optical Images in *Applied Optics and Optical Engineering*, Vol. II, ed. R. Kingslake, pub. Academic Press.

Krabbe, A., Genzel, R., Drapatz, S. & Rotaciuc, V. 1991, *Ap.J.*, **382**, L19.

Krassovsky, V.I., 1972. *Ann. Geophys.*, **28**, 739.

Lacy, J.H., Achtermann, J.M. & Serabyn, E. 1991, *Ap.J.*, **380**, L71.

Lacy, J.H., Townes, C.H., Geballe, T.R. & Hollenbach, D.J. 1980, *Ap.J.*, **241**, 132.

Lebofsky, M.J. & Rieke, G.H. 1987, in *The Galactic Center*, Proceedings of AIP Conference 155, 163.

Liszt, H.S., van der Hulst, J.M., Burton, W.B. & Ondrechen, M.P. 1983, *A.A*, **126**, 341.

Llewellyn, E.J. & Solheim, B.H. 1978, *Planet. Space Sci.*, **26**, 533.

Llewellyn, E.J., Long, B.H. & Solheim, B.H., 1978. *Planet. Space Sci.*, **26**, 525.

Lo, K.Y. 1982, in *The Galactic Center*, Proceedings of AIP Conference 83, 1.

Lo, K.Y. & Claussen, M.J. 1983, *Nature*, **306**, 647.

Lugten,J,B., Genzel, R., Crawford, M.K.& Townes, C.H. 1986,*Ap.J.*, **306**, 691

Lynden–Bell, D. & Rees, M.J. 1971, *M.N.R.A.S.*, **152**, 461.

Marr, J.M., Rudolph, A.L., Pauls, T.A., Wright, M.C.H., Backer, D.C. 1992, *Ap.J.*, **400**, L29.

Mathis, J.S. 1990, *Ann. Rev. Astron. Astrophys.*, **28**, 37.

McCaughrean, M.J. 1988, *Ph.D. thesis*, University of Edinburgh.

Meinel, A.B.,1950. *Ap.J.*, **111**, 555.

Meinel, A.B.,1950. *Ap.J.*, **112**, 120.

Mies, F., 1974, *J. Mol. Spectrosc.*, **53**, 150.

Moreels, G. & Hersé, M., 1977. *Planet. Space Sci.*, **25**, 265.

Mountain, C.M., Leggett, S.K., Selby, M.J., Blackwell, D.E. & Petford, A.D. 1985, *Astron. Astrophysics*, **151**, 399.

Mountain, C.M., Robertson, D.J., Lee, T.J. & Wade, R., 1990. *Instrumentation in Astronomy VII*, ed. D.L., Crawford, (Proc.SPIE, 1235), 25.

Nicolet, M. 1971, in *Mesospherice Models and Related Experiments*, ed. Fiocco, G., Astrphys. Space Sci. Library, pub. Reidel, Dordrecht, **25**, 1

Oliva, E. & Origlia, L., 1992. *A.& A.*, **254**, 466.

Outred, M., 1978, *J. Phys. Chem. Ref. Data*, **7**, 81.

Parmar, P.S., Lacy, J.H. & Achtermann, J.M., 1991, *Ap.J.*, **372**, L25

Peterson, A.W. & Kieffaber, L.M., 1973. *Nature*, **244**, 92.

Rieke, G.H., Telesco, C.M. & Harper, D.A. 1978, *Ap.J.*, **220**, 556

Ramsay, S.K.,Mountain, C.M. & Geballe, T.R. 1992, *M.N.R.A.S.*, **259**, 751.

Ramsay, S.K., Mountain, C.M. & Geballe, T.R. 1993, *in Astronomical Infrared Spectroscopy*, A.S.P. Conference Series, ed. S. Kwok. *in press.*

Rayner, J.T. 1988, *Ph.D. Thesis*, University of Edinburgh

Rayner, J., Hodapp, K., Zinnecker, H. 1991. *in Astrophysics with Infrared Arrays*, A.S.P. Conference Series, ed. R. Elston, 14.

Reid, M.J. 1989, *in The Center of the Galaxy*, IAU Symposium 136, ed. Morris. M., pub. Kluwer Academic Press, p.37.

Saunders, R.H. 1992, *Nature*, **359**, 131.

Schull, J.M. & Hollenbach, D.J. 1978, *Ap.J.*, **220**, 525

Serabyn, E. & Lacy, J.H. 1985, *Ap.J.*, **293**, 445.

Serabyn, E., Güsten, R., Walmsley, C.M., Wink, J.E. & Zylka, R. 1986, *A.A.*, **169**, 85.

Serabyn, E., Lacy, J.H., Townes, C.H. & Bharat, R. 1988, *Ap.J.*, **326**, 171.

Shimazaki, T. & Laird, A.R., 1970. *J. Geophys. Res.*, **75**, 3221.

Sternberg, A. & Dalgarno, A. 1989, *ApJ*, **338**, 197

Stutzki, J., Stacey, G.J., Genzel, R., Harris, A.I., Jaffe, D.T. & Lugten, J.B., 1988, *Ap.J.*, **332**, 379.

Sutton, E.C., Danchi, W.C., Jaminet, P.A. & Masson, C.R. 1990, *Ap.J.*, **348**, 503.

Tanaka, M., Hasegawa, T., Hayahsi, S.S., Brand, P.W.J.L. & Gatley, I. 1989, *Ap.J*, **336**, 207.

Taylor, M.J. & Hapgood, M.A., 1988, *Planet. Space. Sci.*, **36**, 975.

Taylor, M.J., Hapgood, M.A. & Rothwell, P., 1987. *Planet. Space Sci.*, **35**, 413.

Tielens, A.G.G.M & Hollenbach, D., 1985 *Ap.J.*, **291**, 722.

Tollestrup, E.V., Capps, R.W. & Becklin, E.E. 1989, it A.J., 98, 204.

Turner, J., Kirby–Docken, K. & Dalgarno, A. 1977, *Astrophys. J. Suppl.*, **35**, 281

Traub, W.A. & Stier, M.T., 1976. *Appl. Optics*, **15**, 364.

Tuan, F.T., 1976. In *Research in Gravity Waves and Airglow Phenomena*, Rep. AFGL-TR-76-0296, Air Force Geophys. Lab., Bedford, Mass.

Turner, J., Kirby–Docken, K. & Dalgarno, A. 1977, *Ap. J. Suppl.*, **35**, 281.

van Rhijn, P.J, 1921. *Pub. Ast. Lab. Groningen*, **31**, 1.

Wade, R., Geballe, T.R., Krisciunas, K., Gatley, I. & Bird, M.C. 1987, *Ap.J.*, **320**, 570.

Yu, L., Tuan, F.T. & Tai, H.,1980. *J. Geophys. Res.*, **85**, 1297. Zylka, R., Mezger, P.G. & Wink, J.E. 1990, *A. & A.*, **234**, 133.

Mon. Not. R. Astron. Soc. (1992) **259**, 751–760

Non-thermal emission in the atmosphere above Mauna Kea

S. K. Ramsay,[1] C. M. Mountain[2,3] and T. R. Geballe[3]

[1]*Department of Astronomy, University of Edinburgh, Blackford Hill, Edinburgh EH9 3HJ*
[2]*Royal Observatory, Blackford Hill, Edinburgh EH9 3HJ*
[3]*Joint Astronomy Centre, 660 N. A'ohoku Place, University Park, Hilo, Hawaii 96720, USA*

Accepted 1992 June 11. Received 1992 April 24; in original form 1991 December 31

ABSTRACT

We present 1–2.5 μm spectra of the atmosphere above Mauna Kea. The emission in this region is chiefly non-thermal radiation from excited molecules in the upper atmosphere, but above ~2.3 μm the thermal contribution from the telescope and surroundings dominates. The most important component of the non-thermal emission is from vibrational–rotational transitions of the hydroxyl radical, OH^-. In addition to identifying these lines we examine the temporal and spatial variation of this emission, comparing the results with theories of the origin of the variation. We model the effect of this changing background on an hour-long sequence of observations of different exposure times and demonstrate the importance of frequent sky sampling for observations of extended objects.

Key words: line: identification – molecular processes – atmospheric effects – methods: observational – infrared: general.

1 INTRODUCTION

The dominant component of infrared nightglow is the emission from excited vibrational levels of hydroxyl radicals in the upper atmosphere, the OH^- Meinel bands, first identified by Meinel in 1950 and observed in the airglow for vibrational quantum number $v \leq 9$ (Meinel 1950a, b). Variations in the emission intensity are seen over both long and short time-scales. Long-term fluctuations are due to diffusion of the minor atmospheric constituents, which alters the number density of reactants for the key equations that control the OH^- excitation (equations 2 and 3); shorter period variations, of the order of a few minutes to an hour, are connected with the passage of density and temperature perturbations through the upper atmosphere. These gravity waves can be excited by any phenomenon affecting the atmosphere over short time-scales (Hines 1974), e.g. meteorological effects such as thunderstorms (Taylor & Hapgood 1988) or man-made effects. The passage of these waves causes changes in the reaction rates and column densities of the constituents of the upper atmosphere and a corresponding variation in the OH^- emission intensity. Both these phenomena are discussed more fully in Section 4. These systematic variations in the atmospheric background can impose a limit on the sensitivities achievable with the new generation of infrared astronomical instruments. In the spectroscopic, low-background, read-noise-limited regime the major gain in signal-to-noise ratio is obtained by long integrations of up to tens of minutes. Over these time-scales, variation in non-thermal background emission causes an apparent increase in the 'noise' level unless sufficiently frequent sky monitoring is employed.

2 SPECTRA

2.1 Observations

The spectra presented here were observed at the United Kingdom Infrared Telescope (UKIRT) in 1989 November using the cooled infrared spectrometer CGS2. A complete set of filters ranging from 1–2.5 μm was used, with a resolving power of ~250 at the short-wavelength end of the *J* window, increasing to ~800 at the long-wavelength end of the *K* window. Higher resolution observational spectra can be found in Oliva & Origlia (1992); it should be noted that they quote wavelengths in air. Measurements of the sky were made in 'stare' mode and two standard stars (BS 718 and 1543) were observed in 'chop' mode to obtain flux calibration. Further readings were taken with the filters blanked off to allow the dark current, detector offsets and dewar background to be subtracted.

2.2 Data reduction and results

We developed algorithms to describe the signal reaching the detector for both the sky and stars, including effects due to the atmosphere and the instrument. The sky-subtracted star signal is the product of the instrumental transmission, the

atmospheric transmission and the stellar flux distribution, the latter of which we approximate by a blackbody. To obtain the instrumental transmission alone, we first divided a stellar spectrum by a blackbody calculated at the effective temperature of the star and normalized at 2.2 μm, and then used IRTRANS, the atmospheric model of Traub & Stier (1976) to select points at which the atmosphere is transparent. We calculated the instrumental transmission at these points and interpolated between them to find the instrumental transmission for the full window, assuming that it varies slowly with wavelength. Such a transmission curve was produced for the *J*, *H*, and *K* windows.

The sky signal is the product of the sky emission (thermal as well as non-thermal) and the instrumental transmission profile as determined above. Thus, by dividing the raw sky frames by these transmission curves we obtained the spectra of the sky emission. For wavelengths at which the atmospheric transmission is ≤0.75 we set the flux to zero (Figs 1–3).

The wavelength scale was fixed using an argon arc lamp, and identification of the lines achieved by comparison with the wavelengths quoted in Chamberlain (1961). The agreement between our measured wavelengths and those in Chamberlain is better than 0.002 μm, after the latter were corrected to *in vacuo* wavelengths. The wavelengths identified in our spectra, and the transitions involved, are listed in Tables 1–3 for the OH^- lines. Note that the strong line at

Table 1. Hydroxyl lines in the *J* window.

Measured Wavelength (μm)	Transition	v"-v'	Intensity (photon s^{-1} $arcsec^{-2}$)
1.0063	*Q*	9 – 5	3399.6
1.0263	*Q*	4 – 1	2276.3
1.0363	*P*(1)*bl*†	4 – 1	662.0
1.0393	*P*(2)*bl*	4 – 1	766.2
1.0422	*P*(3)*bl*	4 – 1	498.3
1.0466	*P*(4)*bl*	4 – 1	431.0
1.0503	*P*(6)*bl*	4 – 1	357.0
1.0703	*Rbranch*	5 – 2	572.8
1.0824	*Q*	5 – 2	676.9
1.0913	*P*(1)*bl*	5 – 2	572.8
1.0945	*P*(2)*bl*	5 – 2	647.2
1.1002	*P*(3)*bl*	5 – 2	431.4
1.1043	*P*(4)*bl*	5 – 2	357.0
1.1312	*Rbranch*	6 – 3	468.8
1.1430	*Q*	6 – 3	825.7
1.1524	*P*(1)*bl*	6 – 3	572.8
1.1573	*P*(2)*bl*	6 – 3	550.5
1.1636	*P*(3)*bl*	6 – 3	453.7
1.1696	*P*(4)*bl*	6 – 3	364.5
1.1980	*Rbranch*	7 – 4	528.1
1.2092	*Q*	7 – 4	758.7
1.2186	*P*(1)	7 – 4	476.0
1.2254	*P*(2)	7 – 4	550.4
1.2311	*P*(3)	7 – 4	379.3
1.2379	*P*(4)	7 – 4	275.2
1.2660	*Rbranch*	8 – 5	801.1
1.2897	*Q*	8 – 5	1019.1
1.3010	*P*(1)*bl*	8 – 5	591.0
1.3072	*P*(2)*bl*	8 – 5	602.5
1.3128	*P*(3)*bl*	8 – 5	305.6
1.3210	*P*(4)*bl*	8 – 5	*
1.3286	*P*(6)*bl*	8 – 5	*
1.3996	*P*(2)*bl*	9 – 6	*
1.4139	*Rbranch*	2 – 0	*
1.4326	*Q*	2 – 0	*
1.4501	*P*(1)*bl*	2 – 0	937.4
1.4582	*P*(2)*bl*	2 – 0	929.9
1.4662	*P*(3)*bl*	2 – 0	654.7
1.4818	*P*(4)*bl*	2 – 0	825.8
1.4873	*R*(1) – *R*(7)*bl*	3 – 1	1123.3

*Denotes line present in spectrum, but in a region where atmospheric transmission is poor. †*bl* denotes blend of two or more lines.

Table 2. Hydroxyl lines in the *H* window.

Measured Wavelength μm	Transition	v"-v'	Intensity (photon s^{-1} $arcsec^{-2}$)
1.5056	*Q*	3 – 1	*
1.5170	$P_2(1)$	3 – 1	1082.9
1.5235	$P_1(1)$	3 – 1	3884.0
1.5319	*P*(2)*bl*	3 – 1	3556.2
1.5403	*P*(3)*bl*	3 – 1	2572.8
1.5524	*P*(4)*bl*	3 – 1	2259.9
1.5578	*R*(2)*bl*	4 – 2	1708.6
1.5622	*R*(1)*bl*	4 – 2	2483.4
1.5814	*Q*	4 – 2	6715.0
1.5946	$P_2(1)$	4 – 2	1038.2
1.6018	$P_1(1)$	4 – 2	3749.9
1.6059	$P_2(2)$	4 – 2	1634.2
1.6108	$P_1(2)$	4 – 2	4375.7
1.6217	*P*(3)*bl*	4 – 2	2960.2
1.6330	*P*(4)*bl*	4 – 2	1678.8
1.6434	*P*(5)*bl*	4 – 2	1574.5
1.6489	*Rbranch*	5 – 3	2021.5
1.6692	*Q*	5 – 3	6461.6
1.6839	$P_2(1)$	5 – 3	814.7
1.6891	$P_1(1)$	5 – 3	3466.8
1.6944	$P_2(2)$	5 – 3	1366.0
1.6992	$P_1(2)$	5 – 3	3839.3
1.7061	$P_2(3)$	5 – 3	1217.0
1.7101	$P_1(3)$	5 – 3	2706.9
1.7235	*P*(4), *R*(5, 4, 3)*bl*	5 – 3, 6 – 4	1649.0
1.7367	*P*(5), *R*(2)	5 – 3, 6 – 4	1917.2
1.7418	*R*(1)*bl*	6 – 4	1917.2
1.7487	*R*(0)*bl*	6 – 4	576.3
1.7639	*Q*	6 – 4	5821.0
1.7860	*P*(1)*bl*	6 – 4	*

Table 3. Hydroxyl lines in the *K* window.

Measured Wavelength (μm)	Transition	v"-v'	Intensity (photon s^{-1} $arcsec^{-2}$)
1.9635	*R*(2)*bl*	8 – 6	793.4
1.9685	*R*(1)*bl*	8 – 6	1031.6
1.9756	$R_1(0)$	8 – 6	1638.8
1.9845	$R_2(0)$	8 – 6	620.8
2.0003	*Q*	8 – 6	*
2.0218	$P_2(1)$	8 – 6	*
2.0270	$P_1(1)$	8 – 6	2162.8
2.0315	$P_2(2)$	8 – 6	1019.6
2.0400	$P_1(2)$	8 – 6	2710.5
2.0470	$P_2(3)$	8 – 6	*
2.0556	$P_1(3)$	8 – 6	*
2.0722	$P_2(4)$	8 – 6	*
2.0854	$P_2(5)$	8 – 6	573.1
2.0897	$P_1(5)$	8 – 6	811.3
2.1110	*R*(2)*bl*	9 – 7	853.5
2.1183	*R*(1)*bl*	9 – 7	883.2
2.1250	$R_1(0)bl$	9 – 7	1044.0
2.1333	$R_2(0)bl$	9 – 7	496.3
2.1518	*Q*	9 – 7	2201.2
2.1704	$P_2(1)$	9 – 7	409.1
2.1792	$P_1(1)$	9 – 7	1533.7
2.1870	$P_2(2)$	9 – 7	569.2
2.1938	$P_1(2)$	9 – 7	1550.6
2.2042	$P_2(3)$	9 – 7	639.7
2.2108	$P_1(3)$	9 – 7	1461.3
2.2225	$P_2(4)$	9 – 7	492.3
2.2309	$P_1(4)$	9 – 7	1081.7

1.27 μm is not an OH^- line, but the $O_2(^1\Delta_g)$ line produced through chemical reactions at the same altitude as OH^-, 80–100 km. (For a discussion of the chemical production mechanisms, see Howell et al. 1990.)

The results are presented in Figs 1–3. The wavelengths are *in vacuo*. The *Q*, *P* and *R* branches of various vibrational transitions are clearly visible, particularly in the *K* window. The *Q* branch is not forbidden, since OH^- is a heteronuclear radical possessing an electric dipole moment. It is, however, unresolved. Each *P* and *R* transition is a doublet, due to angular momentum splitting, and these are sometimes resolved depending on the resolution of the filters and gratings used. See for example the *K*-window spectrum (Fig. 3): the 9–7 *P*-branch doublets around 2.2 μm are resolved.

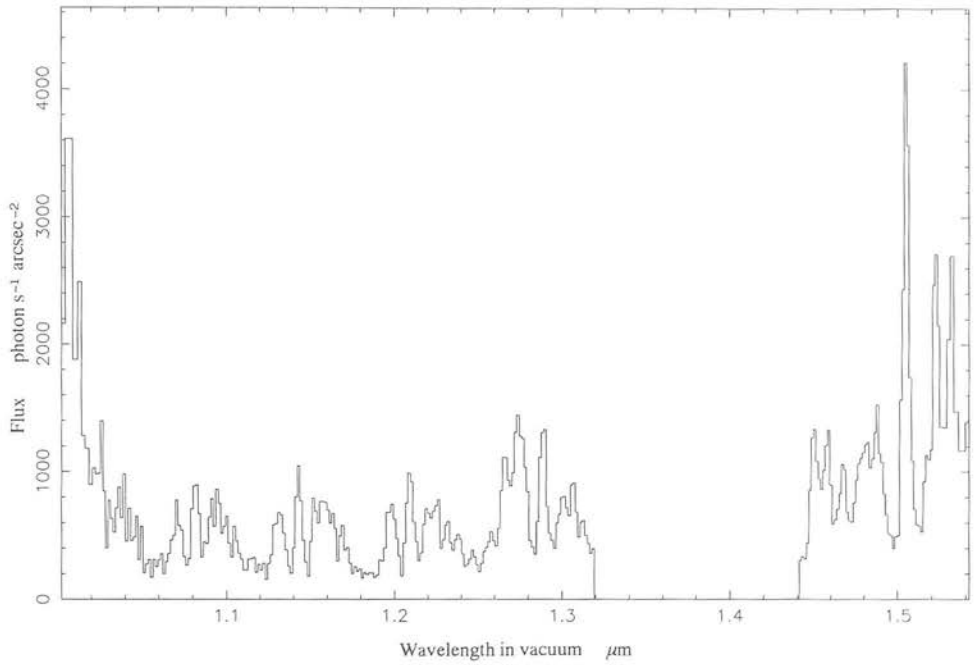


Figure 1. CGS2 spectrum of the OH^- emission lines in the *J* window.

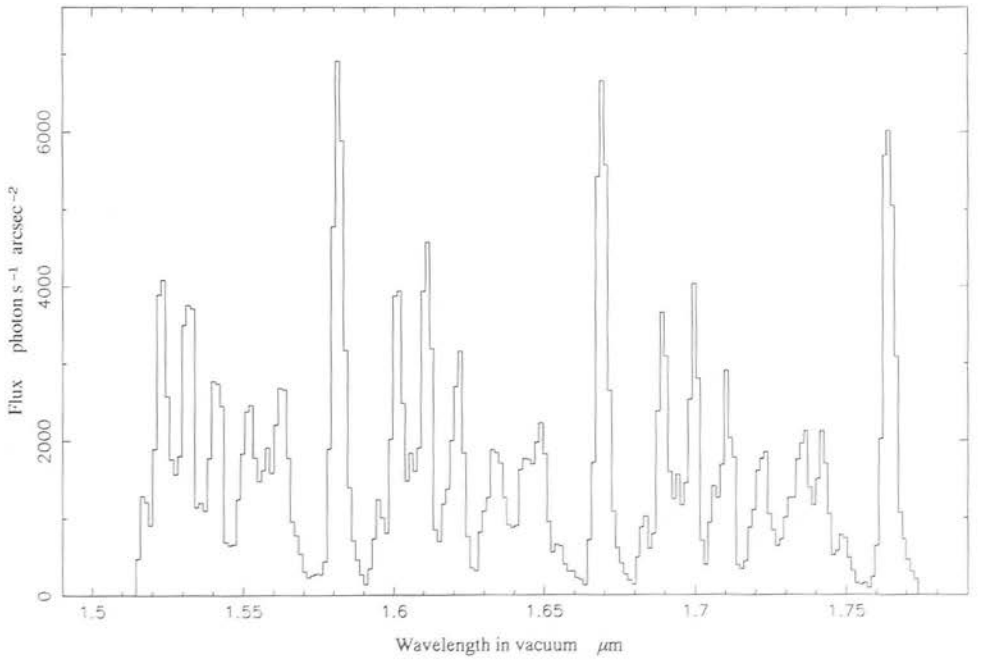


Figure 2. CGS2 spectrum of the OH^- emission lines in the *H* window.

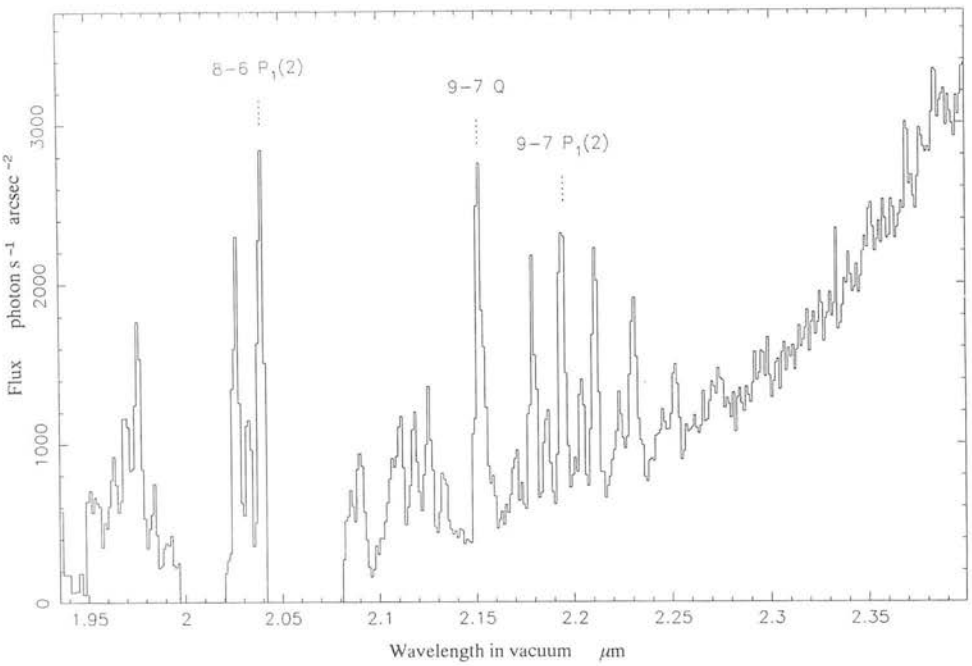


Figure 3. CGS2 spectrum of the OH^- emission lines in the *K* window.

3 TEMPORAL AND SPECTRAL VARIATIONS

3.1 Observations

Information on temporal and spatial variation of the hydroxyl emission was obtained from observations taken with CGS2's successor, CGS4 (Mountain et al. 1990), in 1991 April. CGS4 utilizes an SBRC InSb array of 58×62 pixels. This results in a 90 arcsec long slit on UKIRT, allowing us to examine spatial as well as spectral OH^- variations. The data on the line strengths of the three lines studied, 8–6 $P_1(2)$, $Q9$–7 and 9–7 $P_1(2)$, at 2.0400, 2.1518 and 2.1938 μm respectively, were extracted from a set of 30 observations of a high-redshift galaxy taken on the night of 1991 April 4 between 4.57 and 5.58 am, Hawaiian Standard Time (HST), and a further set of 86 observations taken on 1991 April 5/6 between 9.46 pm and 1.00 am, HST. Our choice of lines enables us to study whether the variations are correlated between vibrational and/or rotational transitions. Each observation consisted of object − sky pairs formed by sequentially offsetting the telescope by ± 30 arcsec along the 3 arcsec wide slit. Each individual observation was fully sampled by stepping the detector four times over two pixels; the total on-chip exposure time was 80 s. The observations were made using a technique for taking multiple samples of the signal during a single exposure which greatly reduces the read noise (Chapman et al. 1990). This gave $40e^-$ read noise per exposure, so our observations were sky-noise limited. Since the object was not detected in either set of observations, we assume that all the line emission at the wavelengths of the OH^- lines comes from that radical. Each observation was dark-subtracted and then flat-fielded using a blackbody at 1073 K to remove pixel-to-pixel gain variations. (The dark current in all cases was negligible, being $< 5e^-s^{-1}$.) We estimate that the flat-fielding was accomplished with an accuracy of 1 part in 10^3 from the statistical variation of 60 exposures, and that any error resulting from this process will be negligible compared with the Poissonian uncertainty in the sky photon flux.

3.2 Data reduction and results

3.2.1 Spatial

To determine the existence of any spatial variations of the OH^- emission on a scale of 90 arcsec we extracted a spatial profile at each of the wavelengths of the lines, at 2.0400, 2.1518 and 2.1938 μm, each frame having first been flat-fielded using the CGS4 calibration unit which illuminates the CGS4 focal plane using a 1073-K blackbody via a gold-coated integrating sphere. From laboratory experiments, we previously determined that the spatial variation in the CGS4 calibration unit is independent of wavelength to about 1 part in 10^4.

Neither the data for the night of April 4 nor those for April 6 showed any evidence of spatial variation along the slit greater than the Poissonian and flat-fielding errors (1 part in 10^3). As will be discussed in Section 4, this is consistent with the gravity wave models of this emission variation. By considering typical parameters for such a disturbance, we find that its effect on the spatial structure of the emission would be dominated by the Poissonian and flat-fielding errors.

3.2.2 Temporal

The spatially flat nature of the OH^- emission during our observations permits us to use the whole slit to find the signal

strength; the values for the OH^- emission strength were obtained by taking the mean signal strength for 16 pixels in the spatial direction. (Bad pixels were ignored to avoid distortion to the mean.) No attempt has been made to calibrate the data. The data reduction system for CGS4 provides signals in 'data numbers per exposure'; to allow us to estimate the Poissonian error, √no. of photoelectrons, correctly, we converted this to electrons by multiplying by the electrons per data number gain for CGS4 (24). During our observations the telescope tracked the object over a range of airmass. The effect of this increasing airmass is to enhance the OH^- signal strength, as the column density along the line of sight is increased. We can compensate for this by dividing the mean signal by the van Rhijn factor, V (van Rhijn 1921), at the zenith angle of the observation, where

$$V=\left[1-\left(\frac{R}{R+h}\right)^2 \sin^2 z\right]^{-1/2}, \qquad (1)$$

and R is the radius of the Earth (6378.17 km), h is the height of the emitting layer (85 km) and z is the zenith angle of the observation. Figs 4 and 5 show the temporal variations for the nights of April 4 and 6 respectively; the line strengths have been normalized to the mean line strength for all the observations and the errors shown on these plots are the Poissonian errors calculated for 16 spectra co-added along the slit. The data from April 4 show strong evidence of a wave-like structure modifying the emission: the amplitude of the variations is a factor of 10 larger than the Poissonian error and is 5 per cent of the total line strength. Simple Fourier analysis of these data yields one significant period of 746 ± 80 s. By ratioing the line strengths for each combination of pairs of lines we see that the short-period, wave-like structure is retained. This effect is most marked in the case of the 8–6 $P_1(2)$ line ratioed with the 9–7 $P_1(2)$ line, but is also present less significantly in the other cases, (9–7) Q/(9–7) $P_1(2)$ and 8–6 $P_1(2)$/(9–7) Q (Fig. 6). From Fig. 4, examination of the temporal behaviour of the emission shows that the lines are varying in phase with each other but that the amplitude of the change in emission intensity is not constant for each line, being greatest for the 8–6 $P_1(2)$ line, followed by the 9–7 Q branch, and least for the 9–7 $P_1(2)$ line: 10, 9 and 8 per cent respectively. The data for April 6, extending over a period of 3 h, also show wave-like structure superimposed on a global decrease of the emission intensity betwen 10 pm and 1 am local time. The amplitudes and periods of the waves are not as regular as for the April 4 data. Due to the small number of cycles present, Fourier analysis of the data set yielded only periods equal to the total observation time, 5600 s, and half that time. We estimate that the range of periods present is from about 5 to 15 min, with the variation in line strength from a few to 10 per cent. The global decrease in the emission strength is 46 ± 5 per cent over the range of the observations. On April 6 also, ratioing the emission for pairs of lines shows that the short-period wave structure is preserved. Again the (8–6) $P_1(2)$/(9–7) $P_1(2)$ ratio demonstrates this effect most dramatically (Fig. 7). The effect is not so marked as on the night of April 4, the change in emission intensity for the three lines being 5 per cent for the 8–6 $P_1(2)$ line, 3 per cent for the 9–7 $P_1(2)$ line and 4 per cent for the 9–7 Q branch (Fig. 5).

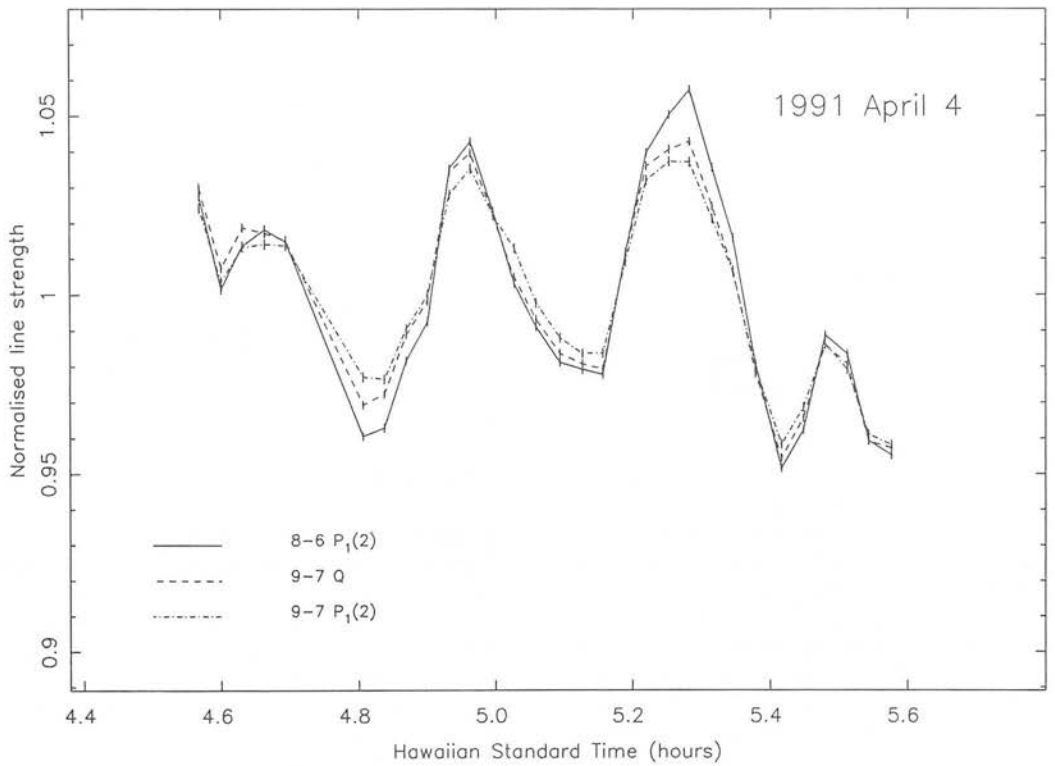


Figure 4. Temporal variation of the OH^- 8–6 $P_1(2)$ line at 2.0400 μm, the OH^- 9–7 Q branch at 2.1518 μm and the OH^- 9–7 $P_1(2)$ line at 2.1938 μm, on 1991 April 4.

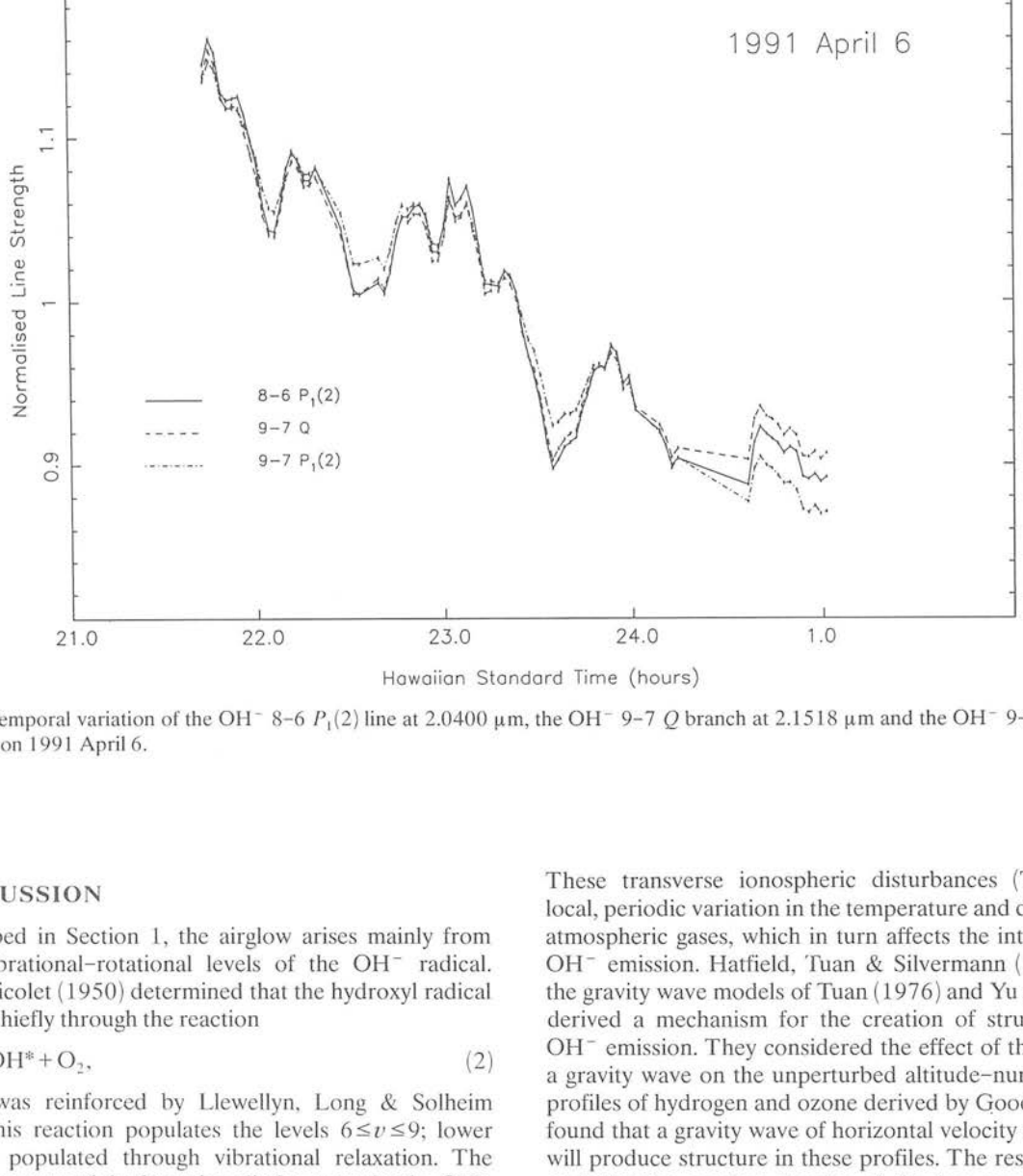


Figure 5. Temporal variation of the OH^- 8–6 $P_1(2)$ line at 2.0400 μm, the OH^- 9–7 Q branch at 2.1518 μm and the OH^- 9–7 $P_1(2)$ line at 2.1938 μm, on 1991 April 6.

4 DISCUSSION

As described in Section 1, the airglow arises mainly from excited vibrational–rotational levels of the OH^- radical. Bates & Nicolet (1950) determined that the hydroxyl radical is excited chiefly through the reaction

$$H + O_3 \rightarrow OH^* + O_2, \quad (2)$$

and this was reinforced by Llewellyn, Long & Solheim (1978). This reaction populates the levels $6 \leq v \leq 9$; lower levels are populated through vibrational relaxation. The emission occurs mainly through radiative cascade; the OH^- has a radiative rate $A \sim 20\ s^{-1}$. The depopulation of the excited states is governed by quenching by other atmospheric components, notably

$$OH^* + O \rightarrow H + O_2, \quad (3)$$

and by vibrational relaxation (Llewellyn et al. 1978). The above quenching reaction is also thought to cause excitation of the oxygen molecule to higher electronic states, which then emit at around 1.27 μm forming another component of the non-thermal background emission, although this is not the dominant method of excitation of O_2. A correlation between emission from these two species has been reported (see e.g. Llewellyn & Solheim 1978).

Short-period variations occur in the nightglow intensity due to the passage of gravity waves through the ionosphere. The existence of such waves has been well documented and modelled (Hines 1960; Tuan 1976; Yu, Tuan & Tai 1980). These transverse ionospheric disturbances (TIDs) cause local, periodic variation in the temperature and density of the atmospheric gases, which in turn affects the intensity of the OH^- emission. Hatfield, Tuan & Silvermann (1981), using the gravity wave models of Tuan (1976) and Yu et al. (1980), derived a mechanism for the creation of structure in the OH^- emission. They considered the effect of the passage of a gravity wave on the unperturbed altitude–number density profiles of hydrogen and ozone derived by Good (1976) and found that a gravity wave of horizontal velocity $v \leq 33\ m\ s^{-1}$ will produce structure in these profiles. The resulting differential in the number density of the key reactants in the chemical excitation of OH^-, H and O_3 at a given altitude results in an enhancement of the OH^- emission at that altitude, as it is governed by the relation

$$I = \eta \sum k[H][O_3], \quad (4)$$

where η is the efficiency of photon emission per OH^- radical formed and k is the reaction rate, $1.5 \times 10^{-12}\ T^{1/2}\ cm^3\ s^{-1}$ (Nicolet 1971). The nature of the structure is a stationary vertical pattern and a transverse horizontal wave, whose phase velocity is the horizontal speed of the gravity wave. The emission peaks at 80–105 km in altitude, which is consistent with observations, including the measurements of 85 km by Hersé et al. (1989) and Taylor, Hapgood & Rothwell (1987). This kind of structure has been imaged using an infrared detecting TV system (Taylor et al. 1987) and photographically (see e.g. Peterson & Kieffaber 1973).

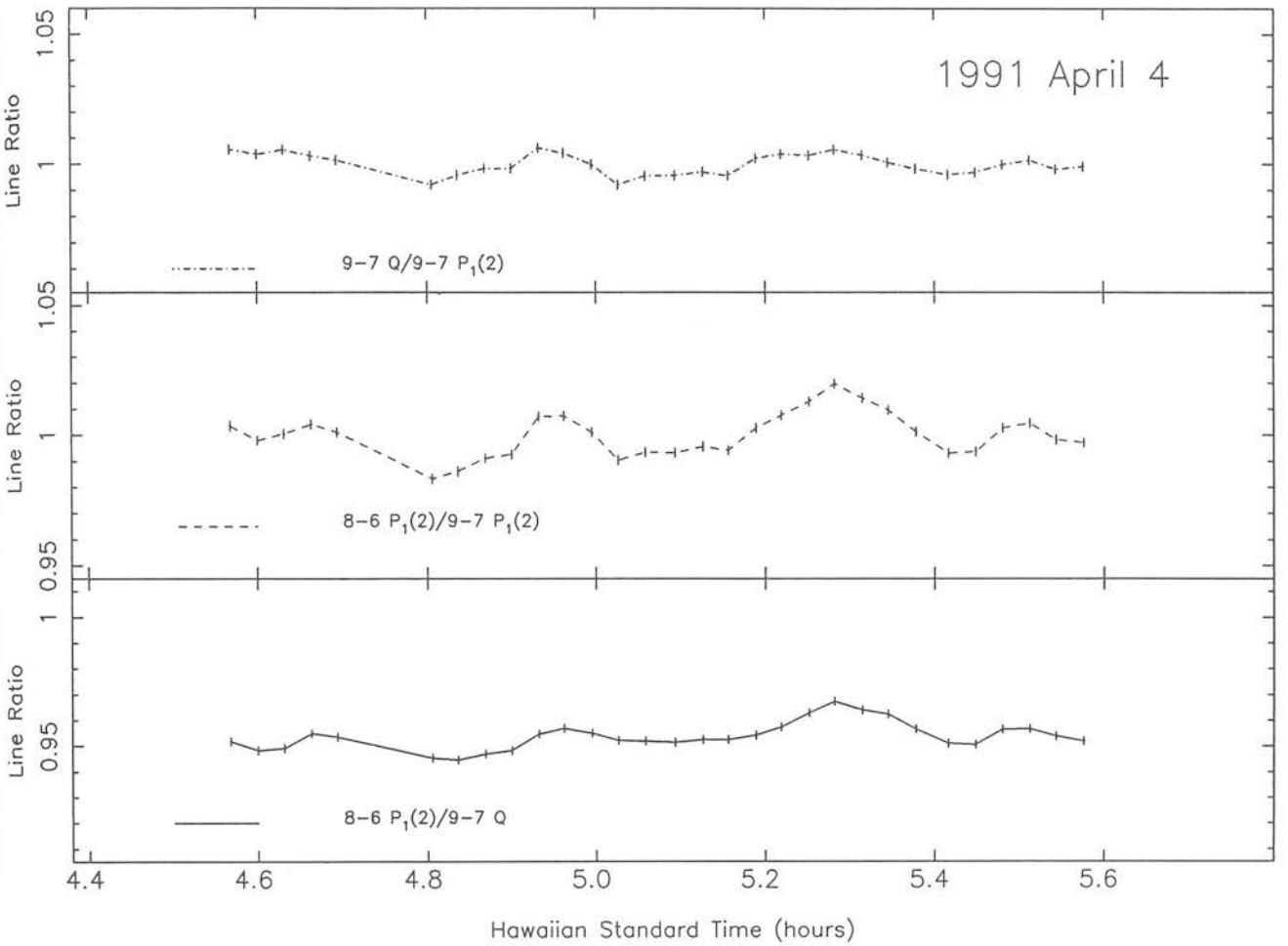


Figure 6. Ratio of fluxes of the 9–7 Q branch to the 9–7 $P_1(2)$ line, the 8–6 $P_1(2)$ line to the 9–7 $P_1(2)$ line and the 8–6 $P_1(2)$ line to the 9–7 Q branch for 1991 April 4.

From examination of the OH^- emission structure, Taylor et al. determined that the wave had a horizontal wavelength of 26 ± 3 km, a horizontal phase velocity of 38 ± 2 m s^{-1} and a period $\tau \sim 11.4$ min for the OH^- structure, and concluded that the structure is due entirely to the passage of the gravity wave of the same period as the emission, $\tau \sim 11.4$ min. Allowing for the angle of propagation, the wavelength of the gravity wave is $\lambda = 12.5$ km, and its phase velocity $v_{ph} = 18.5$ m s^{-1}.

An alternative to this theory of the effect of the passage of a gravity wave on the chemistry of the OH^- emission was proposed by Moreels & Hersé (1977) to explain their images of airglow structure obtained at Pic du Midi. Envisaging the OH^- emissive layer as being of constant thickness but deformed by the passage of the gravity wave to resemble a corrugated sheet, they calculated an enhancement effect analogous to the van Rhijn effect (equation 1) which produces the stripes of bright emission common to all images of OH^- structure.

We did not detect any such spatial variation across CGS4's 90-arcsec slit. Using the parameters derived from our April 4 data, this non-detection can be explained by simple geometry. We consider the variation of the OH^- emission to be a sine wave at a height of 85 km, taking a wavelength of 25 km and a horizontal phase velocity of 40 m s^{-1}, to be consistent with the observations of Taylor et al. (1987). Using these parameters with our measured variation in emission intensity of ~10 per cent, the spatial variation expected across the 90-arcsec slit of CGS4 would be 1 part in 7×10^3, a smaller effect than the combined Poissonian and flat-fielding errors (1 part in 10^3). Our calculations show that this 10 per cent variation in the intensity would be seen on a time-scale of ~5 min, consistent with the results from our 1991 April 4 and 6 data, in which the periods range from 5 to 20 min. The complexity of the structure seen on the night of April 6 could be due to the presence of several gravity wave systems modifying the emission.

The global decrease during the night is predicted by many models of the nightglow emissions, which calculate OH^- emission strength from the nocturnal changes in the altitude–number density profiles of the minor atmospheric constituents (e.g. Gattinger 1969, 1971; Shimazaki & Laird 1970). The dynamical model of Shimazaki & Laird (1970), which also allows for the effects of vertical diffusion of the

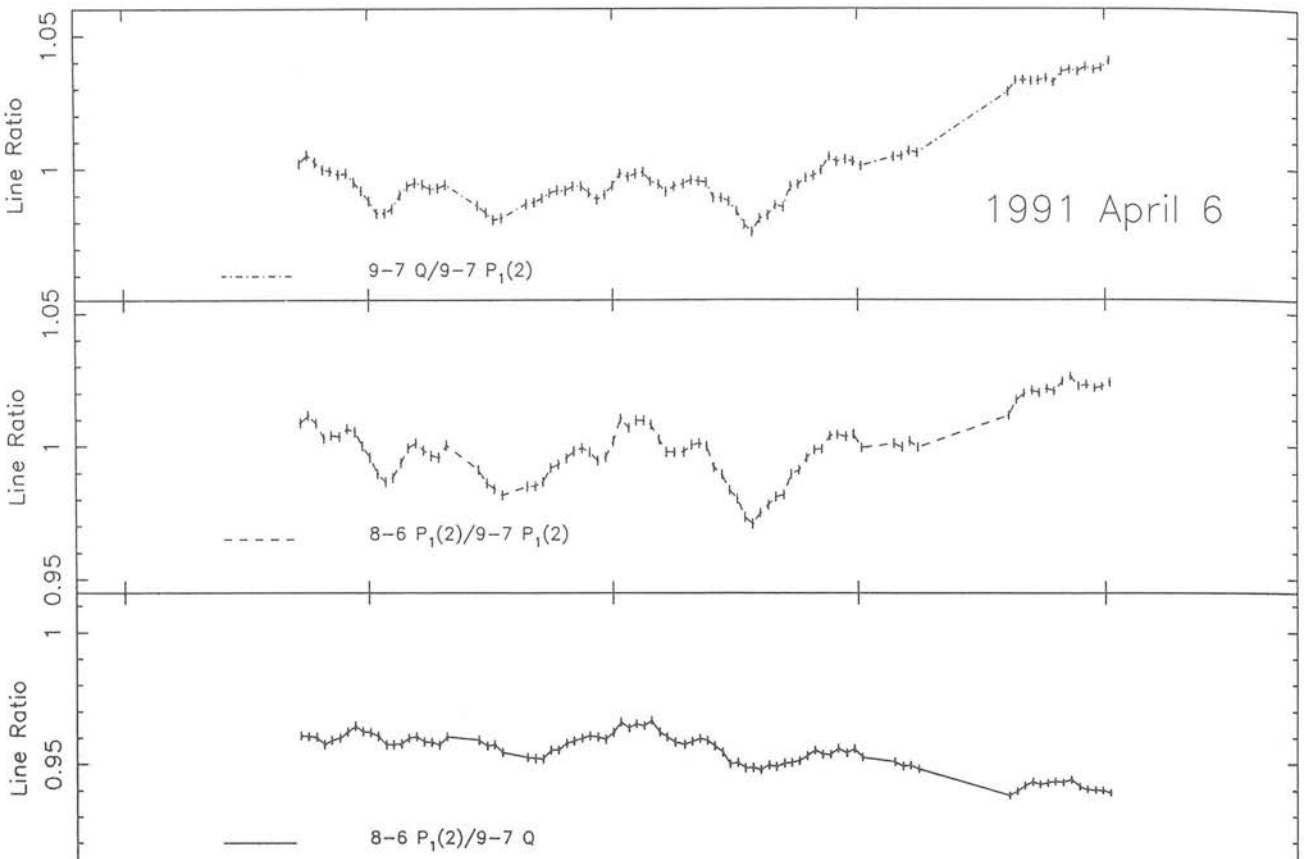


Figure 7. Ratio of fluxes of the 9–7 Q branch to the 9–7 $P_1(2)$ line, the 8–6 $P_1(2)$ line to the 9–7 $P_1(2)$ line, and the 8–6 $P_1(2)$ line to the 9–7 Q branch for 1991 April 6.

constituents, predicts a 50 per cent variation in the OH^- emission rate between the hours of 10 pm and 2 am, which is in excellent agreement with our value of 46 per cent ± 5 for the same period. No such trend was seen in the data for April 4 and, although the observations were taken over a shorter time period, we might reasonably have anticipated seeing such an effect. We propose that the absence of any global decrease is due to the local time of observation (5 to 6 am HST); Shimazaki & Laird (1970) predict that the emission rate is constant at this time.

5 RESIDUAL OH^- SIGNAL IN THE OBSERVATIONS

5.1 The model

To determine an optimum mode for observations of extended objects, we have modelled the residual OH^- signal for an hour of observations. The variation is again considered to be a sine wave with an amplitude of 10 per cent of the OH^- signal strength as seen in the data of April 4. The hour of observations is made up of object − sky pairs; it is assumed that the end of an object observation and the start of a sky observation are coincident. For a given exposure time, the residual signal for an object − sky pair is calculated thus:

$$\mathrm{res-oh}_i = \int_{t=s_{\mathrm{obs}}}^{e_{\mathrm{obs}}} A \sin\frac{2\pi t}{\tau} - \int_{t=s_{\mathrm{sky}}}^{e_{\mathrm{sky}}} A \sin\frac{2\pi t}{\tau}, \qquad (5)$$

where $\mathrm{res-oh}_i$ is the residual OH^- signal on one object − sky pair, s_{obs} and s_{sky} are the start times of the object and sky observations with e_{obs} and e_{sky} the end times, A is the amplitude of the disturbance, here 10 per cent, and τ is the gravity wave period. This is repeated until the sum of the object and sky exposures is one hour, giving a number of pairs associated with the exposure time. The mean residual for this group of observations is then found as follows:

$$\mathrm{mean-res-oh} = \frac{1}{n}\sum_{i=1}^{n} \mathrm{res-oh}_i, \quad \text{where } n \sim 3600/(2\tau). \qquad (6)$$

A further factor which influences this mean residual is the phase of the variation at which the group was started. To allow for this, the mean residual for 500 groups of observations started at random phases was taken and the noise from this mean residual calculated as the standard error of these

residuals. To determine the optimum exposure time, we applied the criterion that the noise from the residual OH^- should be 1 per cent of the photon noise from the OH^- signal itself, and found the exposure times which satisfied this. There are a number of assumptions implicit in this simple model – a more sophisticated model would be required to take into account the fact that the observations become read-noise limited for short exposure times. Also, the residuals for a given phase are assumed to be without error. These assumptions can be justified on the grounds that gross uncertainties in the variation parameters already restrict us to generalizations about the noise behaviour. The results of the model are presented in Fig. 8, which shows residual emission as a function of the exposure time normalized to the period.

5.2 Results

The mathematical form of the residual as a function of exposure time is clearly a sinc^2 function with maxima and minima occurring as a function of the period of the variation. If we knew the precise value of this period we would be able to select an exposure time such that the residual goes to zero. However, this possibility is not open to the observer. Our calculation of the exposure times for which the residual is 1 per cent of the sky noise show that this condition is fulfilled for ranges of exposure time from 0 to 0.1τ and 0.9τ to 1.1τ, where τ is the period of the variation in seconds. At the other extreme, an exposure time of half the period may result in a residual OH^- signal greater than that signal itself, where the maximum, max − oh, is determined from:

$$\mathrm{max} - \mathrm{oh} = \frac{2A\tau}{\pi}, \tag{7}$$

with A and τ defined as in equation (5). Taking a period of 600 s, as seen on April 4, to be typical restricts the observer to using only exposures in the 0 to 0.1τ range; the stronger sky emission lines integrated for 600 s would saturate array detectors currently in use. So we are left with the requirement that the on-chip exposure time, before offsetting to the sky, should be $\leq 0.1\tau$, or in the range 30 to 90 s for variations on scales of 5 to 15 min as seen in our data. It is important to note that this is the simplest model of OH^- emission: the presence of a number of periods in the data, such as in those of April 6, would alter these conclusions. Further work in this field, including characterization of the OH^- emission for a given site and season, will be necessary before we can confidently predict optimum exposure times.

This simple model of the residual OH^- emission indicates that the associated noise due to the variations can dominate even the sky noise in the 1–2.5 μm region. The additional noise contribution provides a fundamental limit to the sensitivity achievable with infrared instruments, and is particularly damaging to observations of faint extended sources.

6 CONCLUSION

We have presented the 1–2.5 μm spectrum of the atmosphere above Mauna Kea, and from it identified the OH^- emission lines, the most significant component of the infrared nightglow. Long-slit spectra from two nights' observations with CGS4 have provided us with information on the temporal and spatial variations of this emission. By examining the behaviour of three of the OH^- lines, we have shown that the spatial variation is negligible over CGS4's 90-arcsec slit, and that the temporal variation over periods of tens of minutes to 3 hours is consistent with theories both of the short-period variation induced by the passage of a gravity wave and of the long-term, diurnal variations in emission strength. This variation was observed to have periods in the range 5 to 15 min, with the variation in line strength lying in

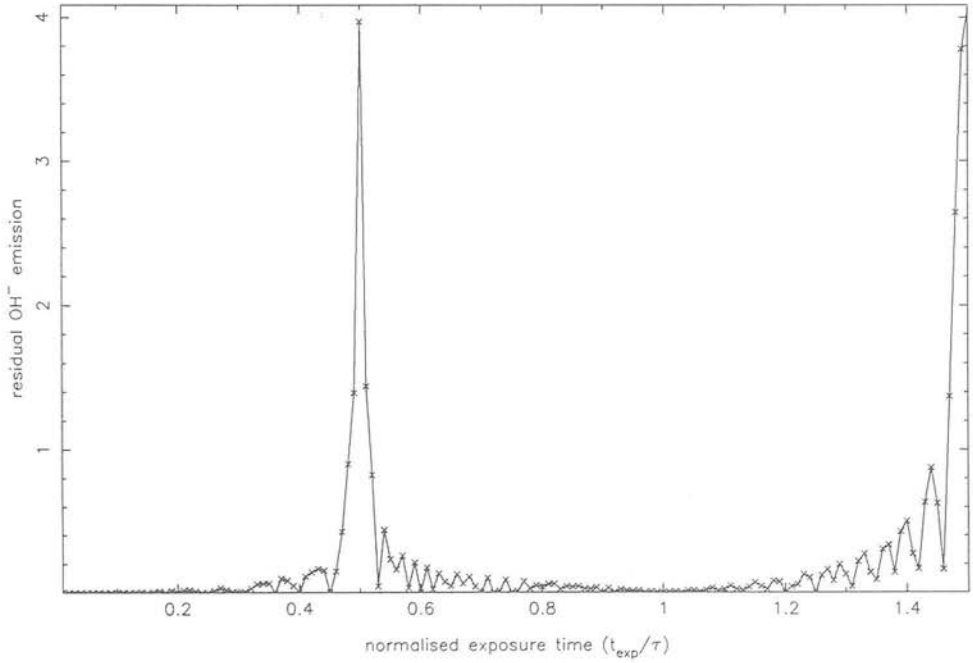


Figure 8. Residual OH^- emission in an hour of observations for a range of exposure times.

the range 3 to 10 per cent. In addition to this, ratioing of the line strengths for each of these lines has demonstrated that the individual vibrational–rotational OH^- lines are not modified identically by the passage of a gravity wave, i.e. the line ratio between any pair of lines is not constant although the general trend, increasing or decreasing flux, is in phase for each line. Modelling of the additional noise from residual OH^- emission has shown that restriction of on-chip exposure times to the range 30 to 90 s is advisable to combat this source of noise. However, these values are presented with the caveats that OH^- variation parameters are poorly determined and that the amplitude of variation and its period strongly affect the residual and the choice of exposure time.

ACKNOWLEDGMENTS

SKR acknowledges the support of an SERC Studentship. The UK Infrared Telescope is operated by the Royal Observatory, Edinburgh on behalf of the UK SERC. We would also like to thank Phil Blanco, Phil Puxley and Martyn Wells for useful discussions during the writing of this paper. We are also indebted to Phil James for his insightful comments. We are grateful to the referee for a number of corrections to the original manuscript and for several helpful comments.

REFERENCES

Bates D. R., Nicolet M., 1950, J. geophys. Res., 55, 301
Chamberlain J. W., 1961, in van Meighem J., ed., Physics of the Aurora and Airglow, International Geophysics Series. Vol. 2, Academic Press, p. 368
Chapman A. R., Beard S. M., Mountain C. M., Pettie D. G., Pickup D. A., Wade R., 1990, in Crawford D. L., ed., Proc. SPIE 1235, Instrumentation in Astronomy VII. SPIE, Bellingham, p. 34
Gattinger R. L., 1969, Ann. Geophys., 25, 825
Gattinger R. L., 1971, in McCormack B. M., ed., The Radiating Atmosphere. Reidel, Dordrecht, p. 51
Good R. E., 1976, Planet. Space Sci., 24, 389
Hatfield R., Tuan T. F., Silvermann S. M., 1981, J. geophys. Res., 86, 2429
Hersé M., Thuillier G., Camman G., Chevassut J.-L., Fehrenbach M., 1989, Appl. Opt., 28, No. 18, 3944
Hines C. O., 1960, Can. J. Phys., 38, 1441
Hines C. O., 1974, Geophysical Monograph No. 18, The Upper Atmosphere in Motion. American Geophysical Union, Washington DC
Howell C. D., Michelangeli D. V., Allen M., Yung Y. L., Thomas R. J., 1990, Planet. Space Sci., 38, 529
Llewellyn E. J., Solheim B. H., 1978, Planet. Space Sci., 26, 533
Llewellyn E. J., Long B. H., Solheim B. H., 1978, Planet. Space Sci., 26, 525
Meinel A. B., 1950a, ApJ, 111, 555
Meinel A. B., 1950b, ApJ, 112, 120
Moreels G., Hersé M., 1977, Planet. Space Sci., 25, 265
Mountain C. M., Robertson D. J., Lee T. J., Wade R., 1990, in Crawford D. L., ed., Proc. SPIE 1235, Instrumentation in Astronomy VII. SPIE, Bellingham, p. 25
Nicolet M., 1971, in Fiocco G., ed., Mesospheric Models and Related Experiments, Astrophys. Space Sci. Library. Vol. 25, Reidel, Dordrecht, p. 1
Oliva E., Origlia L., 1992, A&A, 254, 466
Peterson A. W., Kieffaber L. M., 1973, Nat, 244, 92
Shimazaki T., Laird A. R., 1970, J. geophys. Res., 75, 3221
Taylor M. J., Hapgood M. A., 1988, Planet. Space Sci., 36, No. 10, 975
Taylor M. J., Hapgood M. A., Rothwell P., 1987, Planet. Space Sci., 35, No. 4, 413
Traub W. A., Stier M. T., 1976, Appl. Opt., 15, 364
Tuan F. T., 1976, Research in Gravity Waves and Airglow Phenomena. Rept AFGL-TR-76-0296, Air Force Geophys. Lab., Bedford, MA
van Rhijn P. J., 1921, Publ. Astron. Lab. Groningen, 31, 1
Yu L., Tuan F. T., Tai H., 1980, J. geophys. Res., 85, 1297